

**PAGE MISSING IN
THIS BOOK ONLY
TEXT FLYING THIS
BOOK ONLY**

UNIVERSAL
LIBRARY

OU_148511

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 330/878 E.

Accession No. G 8737

Author Bowman and Bach.

Title Economic analysis and public policy 19

This book should be returned on or before the date last marked below.

30/5/57

11-260

25 NOV 1971

31 OCT 1976

12 DEC 1967

18 DEC 1994

4 MAR 1980

9 SEP 1982

10 FEB 1994

13 DEC 1967

19 OCT 1972

29.10.73

15/10/73

ECONOMIC ANALYSIS
and
PUBLIC POLICY

PRENTICE-HALL ECONOMICS SERIES

E. A. J. JOHNSON, EDITOR

PREDECESSORS OF ADAM SMITH, *by* E. A. J. Johnson.

PLANNED SOCIETY, *edited by* Findlay MacKenzie, Ph. D.

MODERN BANKING, *by* Rollin G. Thomas.

OUR MODERN BANKING AND MONETARY SYSTEM, *by* Rollin G. Thomas.

ELEMENTS OF MODERN ECONOMICS, REVISED, *by* Albert L. Meyers.

MODERN ECONOMIC PROBLEMS, *by* Albert L. Meyers.

MODERN ECONOMICS, ELEMENTS AND PROBLEMS, *by* Albert L. Meyers.

THE THEORY OF PRICES, Two Vols., *by* Arthur W. Marget.

LABOR PROBLEMS AND LABOR LAW, *by* Albion G. Taylor.

A HISTORY OF ECONOMIC THOUGHT, REVISED AND ENLARGED, *by* Erich Roll.

THE ECONOMICS OF WAR, REVISED, *by* Horst Mendershausen.

BUSINESS CYCLES, *by* J. Arthur Estey.

CORPORATE CONCENTRATION AND PUBLIC POLICY, *by* Harry L. Purdy, Martin L. Lindahl, *and* William A. Carter.

ECONOMIC ANALYSIS AND PUBLIC POLICY, *by* Mary Jean Bowman *and* George Leland Bach.

ECONOMIC ANALYSIS and PUBLIC POLICY

An Introduction

by
MARY JEAN BOWMAN, Ph. D.

DEPARTMENT OF ECONOMICS AND SOCIOLOGY
IOWA STATE COLLEGE

and
GEORGE LELAND BACH, Ph. D.

PROFESSOR OF ECONOMICS, CARNEGIE INSTITUTE
OF TECHNOLOGY; FORMERLY OF THE DEPARTMENT
OF ECONOMICS AND SOCIOLOGY,
IOWA STATE COLLEGE

NEW YORK
PRENTICE-HALL, INC.
1946

COPYRIGHT, 1943, BY
PRENTICE-HALL, INC.
70 Fifth Avenue, New York

ALL RIGHTS RESERVED. NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM, BY MIMEOGRAPH OR ANY OTHER MEANS, WITHOUT PERMISSION IN WRITING FROM THE PUBLISHERS.

First PrintingJanuary 1943
Second PrintingJune 1944
Third PrintingOctober 1944
Fourth PrintingFebruary 1945
Fifth PrintingApril 1945
Sixth PrintingApril 1946
Seventh Printing.....May 1946
Elghth PrintingAugust 1946
Ninth PrintingSeptember 1946

PRINTED IN THE UNITED STATES OF AMERICA

Preface

THE AIM of this book on economic analysis and public policy is to be at the same time rigorous in analysis, yet meaningful and teachable in content. Our goal has been so to integrate theoretical and descriptive material as to give greater meaning to both. By applying the theoretical analyses to problems of current public policy, we have sought to avoid a sterility that is unfortunately common in introductory economics courses and to give students practice in analyzing the real-world problems of today and the future. Because we have not evaded currently disputed issues, on occasion we may have failed to give a completely objective picture; as pointed out in the text, this failure is an inevitable result of any attempt to analyze complex policy problems, and we do not consider an occasional misstep of this kind serious so far as the usefulness of the material as a point of departure in the classroom is concerned. At times principles have been stated without the many detailed qualifications that would make them strictly accurate; inclusion of all such refinements on an introductory level would serve far more to confuse students than to improve their grasp of the material concerned. On the other hand, rigorous adherence to the fundamentals of scientific method in social study has been a major purpose—straight thinking is surely a paramount goal in any college course.

The book is so arranged that the material flows logically from one part to another. However, if desired, certain parts may be used as separate units, and some shifting of order is quite feasible. For example, teachers interested in emphasizing theory and minimizing "institutional" material may want to skip from Part IV to Part VII. Others may prefer to omit some of the theoretical material and focus attention on the descriptive parts, which would in that case be less meaningful but nevertheless could be effectively taught. Certain chapters can easily be omitted without disturbing the flow of thought; examples are Chapter 34 on "Legal Regulation of Wages and Hours of Work," and Chapter 42 on "Gold and the Price Level."

In a work of this sort the problem of acknowledgments is an extremely difficult one. Over the past several years many members of the Department of Economics at Iowa State College have participated in teaching the various versions of the book as it has grown into being. Their suggestions and criticisms have proved extremely valuable. To list all those who have made valuable suggestions is impossible, but special acknowledgment is due to a few. Mr. Limen Smythe continually made useful suggestions and wrote some of the material in the early phases of the

project. Mrs. Marjorie Hill Smythe contributed many hours and innumerable penetrating comments. Mr. Lloyd Barber's suggestions on the sections on value and distribution theory have been judicious and ingenious. Dr. Margaret Reid has served throughout as a kind of "provoker" to renewed and better efforts. Outside this college staff, several other economists have generously read and criticized portions of the manuscript, and special recognition is due our editor, Dr. E. A. J. Johnson of New York University, whose keen comments, devastating criticisms, and enthusiastic eulogies spurred us on to make changes that greatly improved the text in clarity and teachability. Lastly, the many classes of elementary students at Iowa State College who have served as willing, and unwilling, guinea pigs should perhaps have the greatest thanks of all. To all these we are happy to express our sincere appreciation. We alone, however, are responsible for the material as it now stands, and of course nothing in the book is to be construed as reflecting the views of the Board of Governors of the Federal Reserve System.

The book was started and largely completed as a joint work of the two authors, while both were teaching at Iowa State College, and cross-fertilization of ideas has been an important part of its development. Authorship is fairly distinctly divided, however, with regard to most of the material included. Miss Bowman is largely responsible for the organization and writing of the whole of what is usually termed "value and distribution theory" and of most of the accompanying "institutional" material, comprising the first seven parts of the book, except that Mr. Bach has shared equally in the writing of several of the early chapters of Part I, and of important sections of Part V. For the entire remainder of the book, beginning with Part VIII, Mr. Bach is largely responsible.

M. J. B.

G. L. B.

Contents

PART I

LAYING THE FOUNDATIONS

CHAPTER

PAGE

1. THE PROBLEM OF ECONOMIZING 3

Production for Consumption: Scales of preferences, planes of living, standards of living; Present American planes of living. Economic Resources and Technology: Natural resources of the United States; Produced resources of the United States; Manpower in the United States; Technology. The Economic System. Place of the Economic System in the Achievement of a Standard of Living. Individuals and the Economic System. Economic Analysis in the Analysis of Social Organization.

2. THE MODERN EXCHANGE ECONOMY 19

The Advantages and Limitations of Specialization: Relation between specialization and exchange; The advantages of specialization and exchange; Limitations of specialization and advantages of diversification. Specialization and Exchange in the American Economy: Factors encouraging specialization in the American economy; Occupational specialization in America today; Regional specialization in America today; Specialization in the production of a single consumers' good. The Functions of "Middlemen" in a Complex Exchange Economy: Exchange; Information; Physical supply; Auxiliary functions; The quantitative importance of middlemen; Are there too many middlemen? A Simplified Model of a Free Price System: How the free price system registers consumers' preferences; How the free price system gets desired goods produced; How the free price system distributes the products of economic activity; The circular flow of economic activity. The Operation of the American Economy: Displacement of individual by group action; Government in the American economy. The Method of Economic Theory.

3. CONSUMER DEMAND 43

The Meaning of the Term "Demand": Individual demand; Aggregate demand. Graphic Representation of Demand. Changes in Demand: Graphic presentation of a change in demand; Factors causing a change in demand. The Downward Slope of Demand Curves: The downward slopes of individual demand curves; The downward slope of aggregate demand curves. Elasticity of Demand: The meaning of elasticity of demand; Demand elasticity and graphical analysis; The significance of demand elasticity; Factors affecting demand elasticity. Interrelations of Demands for Different Things. Producer Response to Consumer Demands.

4. THE UNIT OF BUSINESS ENTERPRISE	61
The Single Proprietorship: Legal form; Advantages; Disadvantages; Decreasing importance. The Partnership: Legal form; Advantages; Disadvantages; Extent of the partnership form of enterprise. The Corporation: Legal form; The financing of corporate enterprise; Advantages of the corporate form of enterprise; Disadvantages of the corporate form of enterprise; Prevalence of the corporation; The corporation as a device for concentrating control over large investments. The Co-operative.	
5. SIZE AND CONCENTRATION IN AMERICAN ENTERPRISE	77
Some Basic Economic Concepts. Size of Enterprise in the United States: Trends in the size of manufacturing establishments; Industrial consolidations; Trends in the size of enterprises in retailing; Trends in the size of agricultural enterprise. Factors Influencing Size of Enterprise: Technology and the size of business enterprise; Management and the size of enterprise; Market opportunities and the size of business enterprise; Financing and financial manipulation. The Number of Firms in an Industry: Government-operated enterprise and regulated public utilities; Degrees of concentration in manufacturing; Lack of concentration in agriculture; Degrees of concentration in mining, construction, and retailing and consumer services. Relations Between Leadership of the Firm and Diversification of Products. The Causes of Product Diversification: Response to diversified customer demands; Efficient utilization of resources; The quest for security; Diversification as a corollary of integration; Attaining of dominant and controlling market positions. Summary and Conclusions.	
6. PHYSICAL INPUT-OUTPUT RELATIONS IN A FIRM	106
Long- and Short-Run Planning. Short-Run Combinations of Productive Agents: "Total," "marginal," and "average" product; Input-output data for a hypothetical firm; Graphic presentation of input-output data. Long-Run Combinations of Productive Agents.	
7. COSTS OF PRODUCTION	117
The Meaning of "Costs": The "explicit" and "implicit" components of costs; Costs and profits as viewed by a manufacturing corporation; Costs and profits as viewed by an independent farmer. A Short-Run View of Costs: The short-run distinction between fixed and variable costs; Total, fixed, and variable costs per unit of output; Short-run output-cost curves; The decision to operate or to shut down in the short run. A Long-Run View of Costs: The "optimum scale of enterprise"; "Economies" and "diseconomies" of scale of enterprise; Long-run output-cost curves of the firm; Divergences from optimum scale; Total costs and the entry or withdrawal of firms.	
8. THE FIRM IN THE MARKET	133
Classification of Types of Market Positions: Pure competition in selling and buying; Simple monopoly in selling, simple monopsony in buying; Monopolistic competition in selling, monopsonistic competition in buying; Oligopoly in selling, oligopsony in buying. Freedom and Restrictions of Entry of Firms in the Long Run.	

CONTENTS

ix

CHAPTER

PAGE

9. ECONOMIC THEORY AND SOCIAL POLICY 140

Some Frequently Accepted "Ends" of Social Policy: Allocation of resources in accordance with consumer preferences; Freedom in the choice of a job or business; Equity in the distribution of income; Progress in raising planes of living; Stability and security. The Role of Economics: The economist and the clarification of social ends; The economist and the selection and interpretation of facts; The economist and the selection of means; The institutional framework of economic analysis.

PART II

EQUILIBRIUM OF THE FIRM

10. PRELIMINARY ANALYSIS OF SHORT-RUN POLICIES OF THE FIRM 151

The Production Policies of Purely Competitive Firms: The maximum profit (or minimum loss) level of production in a hypothetical purely competitive enterprise; The decision to operate or shut down in the short run; Total unit costs and the position of the firm; Variable unit costs and the position of the firm; Graphic representation of profits and losses of a purely competitive firm. The Policies of Firms Buying Purely Competitively and Selling Monopolistically: Production and price adjustments of a monopolistic firm; Graphic representation of profits and losses. The Policies of Monopsonistic Firms: A hypothetical monopsonistic buyer selling his product in a purely competitive market; A monopsonistic buyer selling monopolistically. Monopolistic Sellers Face Monopsonistic Buyers. Price Discrimination.

11. INTRODUCTION TO MARGINAL ADJUSTMENTS 170

Defining Some Marginal Concepts. Marginal Costs and Average Expenditures of Purely Competitive and of Monopsonistic Buyers: Purely competitive buying; Monopsonistic buying; Graphic representation of the relation between marginal cost and average expenditure; Average expenditure as a measure of the "worth" of a unit of input in other uses. Marginal Revenue and the Sale Value of the Marginal Products of Purely Competitive and of Monopolistic Sellers: Marginal revenue and the elasticity of an average revenue schedule; Marginal revenue and the sale value of the marginal product; Sale value of the marginal product as a measure of the "worth" of the marginal product to consumers.

12. MARGINAL ANALYSIS OF THE MOST PROFITABLE LEVEL OF PRODUCTION 181

Adjustments in Purely Competitive Firms: Marginal analysis of the most profitable level of production; Evaluation of the level of production in a purely competitive firm. Adjustments in Firms That Are Not Purely Competitive in Both Buying and Selling: A hypothetical firm that buys purely competitively and sells monopolistically; A hypothetical firm that buys monopsonistically and sells purely competitively; A hypothetical firm that buys monopsonistically and sells monopolistically. Graphic Representation of Marginal Input Adjustments.

13.	ANALYSIS OF MARGINAL ADJUSTMENTS IN TERMS OF OUTPUTS	193
	Examination of Concepts: Deriving figures for marginal unit cost and marginal unit revenue from discontinuous production data; The relation between marginal unit revenue and average revenue; Failure of marginal output analysis to discriminate between pure competition and monopsony in buying. Finding the Most Profitable (or Least Unprofitable) Output by Comparing Marginal Unit Revenues and Marginal Unit Costs. The Relations Between Marginal Unit Costs and Other Cost Data: The relation between marginal unit costs and variable unit costs; The relation between marginal unit cost and total unit cost. Cost-Output Curve Analysis and the Short-Run Adjustments of the Firm: Presentation of cost and revenue curves; Maximum profit (or minimum loss) output and the least-cost-combination.	

PART III

PURELY COMPETITIVE MARKET ADJUSTMENTS

14.	PURE COMPETITION AND THE LAW OF SUPPLY AND DEMAND	207
	Adjustment Toward Competitive Equilibrium. Effects of Changes in Demand or Supply on Equilibrium Price and Quantity: Changes in demand; Changes in supply: Changes in both demand and supply; Price and quantity variability. The Difficulties of Price Fixing. Periods of Supply.	
15.	THE FIRM AND SHORT-RUN MARKET ADJUSTMENTS . .	216
	The Firm and Purely Competitive Supply: The supply schedule of an individual firm; Aggregate short-run supply; Preliminary statement of the pricing of the products of purely competitive firms. The Firm and Purely Competitive Demand for Productive Services: The Preliminary assumption that a unit of output contains one productive agent only; The demand schedule of an individual firm; Aggregate short-run demand. A Note on Partial Equilibrium Analysis.	
16.	LONG-RUN PRICE AND OUTPUT ADJUSTMENTS	225
	The Process of Adjustment Toward Long-Run Equilibrium: Response to profit expectations; Looking ahead at costs and revenues; Illustration of response to an increase in demand; The long-run supply schedule. Relation Between Costs and Long-Run Supply: Survival of the fittest and long-run uniformity of lowest total unit costs; The long-run aggregate supply curve—a long-run cost curve of the industry; Relation between costs to a firm and the size of an industry. A Re-Examination of Influences Determining Supplies of Products: Elasticity of supply of a productive service and supply of the product; Influences determining the supply of a productive service to an industry; Summary of the factors influencing the supply of a product.	

CHAPTER

PAGE

17.	LONG-RUN DEMANDS FOR PRODUCTIVE SERVICES . . .	239
	<p>Derived Demand for a Productive Service: A long-run demand, illustration of an adjustment to an increase in the supply of a productive agent; Relation between elasticities of demands for products and the services of a productive agent. Interrelated Demands of a Firm for Services of Different Productive Agents: Characteristics of the most profitable combination of a number of productive agents in a firm; Complementary and substitutive relations between productive agents; Adjustment of a firm to a change in price of one among a number of productive agents. Summary of Factors Influencing Prices and Aggregate Demands for the Services of a Productive Agent. Appendix to Chapter 17, Advanced Analysis of Interdependent Demands for Productive Services: Input adjustments assuming two agents variable; Interdependence in the pricing of productive services.</p>	
18.	JOINT PRODUCTS	252
	<p>Characteristics of "Joint Products": "Joint products" and "joint costs"; Complementary and substitutive production relations between joint products; Pure competition and the position of joint product firms. The Output and Pricing of Joint Products That Are Perfectly Complementary: Output policy of the firm; Short-run supply of "perfectly complementary" joint products; Demand schedules for joint products; The short-run pricing of "perfectly complementary" joint products; Long-run adjustments. The Output and Pricing of Joint Products the Proportions of Which Can Be Varied: Output policy of the firm; Adjustment of a firm to a change in the price of one among several joint products; The supplies and prices of joint products; Long-run adjustments. Significance of Joint Products in the Use of Single Product Analysis. Appendix to Chapter 18, Advanced Analysis of Output Adjustments of Joint Products the Proportions of Which Can Be Varied.</p>	
19.	PURELY COMPETITIVE GENERAL EQUILIBRIUM	267
	<p>The Characteristics of General Equilibrium in a Purely Competitive Economy. Practical Usefulness of the General Equilibrium Concept.</p>	
20.	THE CONSUMER LOOKS AT PURE COMPETITION . . .	271
	<p>What Are Consumer Preferences? The division of consumer votes; The expression of individual preferences. Consumer Preferences and the Revenues and Costs of Firms: The sale values of marginal products as indexes of consumer preferences; Opportunity costs and sale values of marginal products. Conclusion: The Allocation of Resources Under Pure Competition: Short-run allocation of resources under pure competition; Long-run allocation of resources under pure competition.</p>	

PART IV

MONOPOLISTIC AND MONOPSONISTIC MARKETS

21.	SIMPLE MONOPOLY AND SIMPLE MONOPSONY	283
	<p>Simple Monopoly: Exclusion of new firms as the basic foundation for simple monopoly; Relation between "decreasing costs" and simple</p>	

CHAPTER

PAGE

21. SIMPLE MONOPOLY AND SIMPLE MONOPSONY (*Cont.*):

monopoly; Short-run maximum profit pricing of products of a monopoly; The effects of long-run considerations on the price policies of a monopolist; Errors in estimating maximum profit positions; Scale of enterprise; Profits in the long run. Simple Monopsony: The occurrence of simple monopsony; Short-run maximum-profit pricing of productive services by a monopsonist; The effects of long-run considerations on the price policies of a monopsonist.

22. MONOPOLISTIC COMPETITION AND MONOPSONISTIC COMPETITION 294

Monopolistic Competition: The bases of product differentiation; Short-run price adjustments; Quality competition; Long-run adjustments. Monopsonistic Competition: The basis of "outlet differentiation"; Pricing of the services of material agents of production; Some peculiarities of the pricing of services of labor; Long-run adjustments. Appendix to Chapter 22: Graphic representation of long-run adjustments under monopolistic competition; Long-run elimination of profits with freedom of entry to a "production area."

23. OLIGOPOLY AND OLIGOPSONY 303

Price Competition: "Nonaggressive" price competition among oligopolists; Oligopoly collaboration in price and output policies; Non-aggressive price competition among oligopsonists; Oligopsony collaboration in price and input policies; Crudeness of approximations to maximum profit marginal adjustments in a firm; The shunting of competition into non-price channels or indirect price competition. Long-Run Adjustments: Long-run adjustments with freedom of entry; Restrictions on entry. Appendix to Chapter 23: Graphic presentation of aggressive versus nonaggressive price competition.

24. NON-PRICE COMPETITION AND PRICE-CONTROL EVASIONS AMONG SELLERS 313

Quality Competition: The problem of measuring quality; Range and emphasis in quality competition. Demand Creation: Increasing the demand for the firm's product by advertising; The meaning of "selling costs"; Sellers' difficulties in estimating the effects of advertising; Competitive demand creation; Advertising versus price-cutting and quality competition in business policy. Escape Devices in Lieu of Open Price Competition: Conflicts of interest as revealed under the National Recovery Act; Devices to evade price control schemes as revealed in NRA regulations; Analogous evasions of price ceilings.

25. A PHILOSOPHIC INTRUSION: INDIVIDUALISM AND THE ECONOMY 325

Monopolistic and Monopsonistic Interferences with Resource Use: Short-run allocation of productive agents; Long-run under- and over-development of an industry. The Relation Between Payments to Productive Agents and "Economic Contribution": Payments and "economic contributions" in a purely competitive economy; Monopolistic enterprise and payments to owners of productive agents; Monopsonistic markets and payments to owners of productive agents. The Effects of Quality Competition: The significance of shifting emphasis from price to quality competition; Gaps and wasteful duplication; Paying for variety; Errors in sellers' judgments of consumer

25. A PHILOSOPHIC INTRUSION: INDIVIDUALISM AND THE ECONOMY (*Cont.*):

demands. The Effects of Demand Creation: The effects of advertising on consumer selections in the market; Advertising as a basis of monopolistic power; Shifting allocation of resources between production and advertising; Advertising as a way of meeting uncertainties in modern markets. Product Diversification and the Character of Competition: Response to consumers' preferences in joint-product industries approximating the model of pure competition; The use of some products to increase sales of others; Product diversification as a basis of strength in conducting tactics of exclusion; Product diversification and product differentiation. Entrepreneurial Mobility and Economic Opportunity: Ignorance of alternatives; Entrepreneurial preferences; Costs and difficulties of readjustment; Size of initial investment; Restrictions on entry. Labor Mobility and Economic Opportunity: Ignorance; Preferences; Differences in abilities; Costs of shifting employments; Discrimination against particular groups of people; Exclusive tactics of organized labor groups; Labor mobility, resource allocation, and economic opportunity. Conclusion: The philosophy of individualism and the "atomistic" view of society; The strength and ruthlessness of "rugged" individualism; The contrasting extreme in a philosophy of totalitarianism; Evolving concepts of individualism in American Democracy.

PART V

MARKET CONTROL AND PUBLIC POLICY IN AMERICAN ENTERPRISE

26. FOUNDATIONS OF POWER 351

Close-Knit Federations: The corporation in the organizational framework; The trusts; The holding companies; Business consolidations. Loose-Knit Federations Privately Organized: Informal agreements; The "pools" of the nineteenth century; Trade associations as the bases of modern "pools" or "cartels"; Cooperative selling agencies; Interlocking directorates. Publicly Fostered Cartelization: Codes of "fair competition" and the National Industrial Recovery Act; Recent cartelization under federal auspices. Controlling and Creating Bottlenecks: Controlling transportation facilities; Controlling marketing channels; Controlling raw material and power sources; Patents; Controlling the labor supply. Discriminatory Price Cutting: Some examples in the history of the oil-refining industry; Some examples in the meat-packing industry. Manipulations of Laws Concerning Standards: Unreasonable licensing requirements: an example from the plumbing industry; Unreasonable requirements concerning materials or conditions of sale; Public requirement that sellers use a privately owned grademark. Conclusion.

27. SOME MONOPOLISTIC PRICE POLICIES 380

Extreme Price Policies: Examples of extreme price policies of the big trusts of the nineteenth century; An example of recent high-price policy; Speculative manipulation of prices by outside financiers. Price Stabilization as a Business Policy: Incentives to collaborative price stabilization in high-fixed-cost industries; Price leadership and stabilization of steel prices; Basing-point pricing; The extent of price stabilization.

27. SOME MONOPOLISTIC PRICE POLICIES (*Cont.*):

Recent Price Policy Under Governmentally Sponsored Cartelization: Pricing under the Agricultural Adjustment Administration; The pricing of fluid milk; The pricing of bituminous coal. A Historical Perspective on Price Policies: Privately controlled pricing programs; Recent government support of collaborative pricing.

28. PUBLIC POLICY ATTACKING RESTRAINTS OF TRADE IN BUSINESS 397

Common Law and Restraints of Trade: Contracts in restraint of trade; Integrations; Summary of the position of the common law. Federal "Antitrust" Laws: Provisions and interpretations of the Sherman Act; Legislation of 1914. Enforcement of the Antitrust Laws; The new attack on restraints of trade since 1938: staff and scope; The relation between enforcement and uncertainty; Penalties under the Sherman Act. Can an Antitrust Program Preserve Competition? The meaning of "competition" in the application of the antitrust laws; Oligopoly and restraint of trade in the antitrust program; The problem of size; The problem of bottlenecks; The income problem; Antitrust in a changing scene.

29. PUBLIC UTILITY REGULATION 415

The Machinery of Regulation. Privileges Granted Public Utility Companies: The exclusive right to operate; The right of eminent domain. Regulation of Quality and Quantity of Sales. Rate Regulation: The problem of a "fair return"; The problem of a "fair value" of investment; Estimating operating expenses; The importance of demand elasticity in the setting of rates; Joint products and differential rates; Evaluation of public utility rate regulation. Regulation of Utility Holding Companies. Government Regulation and Government Ownership. The Public Utility Approach and Antitrust Policy: The regulation of marketing bottlenecks; The public utility approach to patents; Contrast between the "public utility approach" and the NIRA approach to the control of industry. Appendix to Chapter 29.

PART VI

LABOR ORGANIZATION AND LABOR MARKETS

30. THE HISTORY AND PHILOSOPHY OF TRADE-UNIONISM . 439

Trade-Union Purposes and Social Philosophies: Trade-unionism and the collective management of scarce job opportunities; Trade-unions viewed as instruments for the attaining of a new social order. History of Trade-Unionism: The development of trade-unionism in Great Britain; Checks to the development of trade-unionism in the United States; History of trade-unionism in the United States; Trade-unionism in the current American scene; A comparison of trade-unionism in England and the United States.

31. PARTICIPATION OF ORGANIZED LABOR IN THE CONTROL OF ECONOMIC ACTIVITY 456

Attitudes Toward Collective Bargaining: Attitudes and arguments of workers; Attitudes and arguments of employers. The Contents of

31. PARTICIPATION OF ORGANIZED LABOR IN THE CONTROL OF ECONOMIC ACTIVITY (*Cont.*):

Collective Agreements: Provisions concerning employment and job protection; Wage and hour provisions; Provisions for enforcement. Labor Disputes: Union policies and weapons in collective bargaining; Employer policies and weapons in collective bargaining. Looking Forward.

32. THE ECONOMICS OF TRADE-UNIONISM 467

Economic Effects of Exclusions from "Scarce Job Opportunities": Restrictions on entry into the union; Seniority rules; Jurisdictional disputes between rival unions; Resistance to displacement of union workers by machines or by unskilled men. Economic Effects of Wage and Hour Policies in Trade-Union Bargaining: Effects of union wage agreements where buyers of labor services are purely competitive; Effects of union wage agreements where buyers of labor services are monopsonistic or oligopsonistic; The effects of reducing hours worked per week on trade-union positions in wage-bargaining; Factors influencing trade-unions in their wage and hour policies. Evaluation of Some of the Effects of Policies of Job-Conscious Unionism: Trade-unions and the allocation of resources; Realization of workers' preferences; Equity in income distribution; Security and stability; Raising planes of living.

33. GOVERNMENT AND TRADE-UNIONS 488

The Legal Powers of Labor and Employers: Traditional principles in the law; The right to bargain collectively; The law and the weapons of labor; The law and the weapons of employers. The Adjustment of Labor Disputes. Trade-Unionism and Restraint of Trade: Recent Court decisions in antitrust labor cases; The position taken by the Antitrust Division; Evaluation of arguments of the opponents of the Antitrust Division's position in labor cases; Conclusion.

34. LEGAL REGULATION OF WAGES AND HOURS OF WORK . 506

The History of Minimum Wage Legislation: State laws; Interstate pacts; Federal action. Analysis and Evaluation of the Effects of Minimum Wage Regulation: Some effects of minimum wage regulations under purely competitive conditions; Minimum wage regulation in an employer-monopsonized labor market; Minimum wage regulations and management efficiency; Consideration of the argument that "a minimum wage tends to become a maximum"; Evaluation of the argument that higher wages may help to bring prosperity. Some Practical Difficulties in Setting Minimum Wages: Determination of the base rate; The problem of coverage. History of Legislation Controlling Hours of Work.

PART VII

THE DISTRIBUTION OF INCOME

35. THE DISTRIBUTION OF PERSONAL INCOMES 521

Types of Income: Money income; Real income; Psychic income. National Income: National real income; National money income; National money income as an index of national real income. Estimates

35. THE DISTRIBUTION OF PERSONAL INCOMES (*Cont.*):
of Income Distribution in the United States: Average per capita income; Income distribution by family units. Income distribution by occupation; Income distribution by area; Age, sex, and race differences in income levels; Income redistribution. Relation Between Income Distribution and Resource Allocation: Income distribution as the determinant of dollar votes; Expenditure patterns by income groups; Effects of unequal income distribution on the allocation of resources; Effects of consumer spending on the distribution of income.

36. THE SOURCES OF PERSONAL MONEY INCOMES 537
The Flow of Money Through the Firm: The path of a hypothetical dollar from consumer spending to consumer incomes; The functional division of money flows through the firm; Plan of study. Property Income and Purely Competitive General Equilibrium: Net income from material agents of production; "Rent" and "interest" in purely competitive general equilibrium; "Wages" and "interest" in purely competitive general equilibrium; The rate of interest and the allocation of resources. "Pure Profits": The "profit motive" and elimination of "pure profits" in long-run adjustments; Widespread fluctuations in general business activity; Changes in demand for one product relative to others; Changes in supplies of productive services to an industry; Innovations; "Risk" versus "uncertainty." Monopolistic and Monopsonistic Elements in the Distribution of Incomes: The wages of labor; Rates of interest; Profits of exclusive position. Sources of Personal Money Incomes in the United States: Comparison of statistics of income sources with income concepts in economic analysis; Principal sources of income in the United States; "Capital gains and losses"; Income sources by size of income.

37. INCOME DISTRIBUTION AND PUBLIC POLICY 561
"Economic Contribution" and the Distribution of Income: "Economic contributions" and purely competitive general equilibrium; "Economic contribution" and "uncertainty"; "Economic contribution" and economic opportunity; The criterion of economic contribution and the allocation of resources. Need and Equality: Comparison of an important socialist ideology with attitudes in the United States; Arguments commonly presented for and against equalization of income distribution. Public Policy and Income Redistribution: The measurement of inequality of income; Redistribution and the collection of public revenues; Redistribution and public expenditures. Controlling Income Distribution at the Source: Control within a private enterprise economy; Income distribution in a socialistically organized economy.

PART VIII

MONEY, CREDIT, AND PRICES

38. INTRODUCTION TO MONEY AND CREDIT 579
Money in a Complex Society. "Money" and "Credit." Types of Money. Types of Credit

CONTENTS

xvii

CHAPTER

PAGE

39. THE BANKS AND THE QUANTITY OF CREDIT 586
Middlemen Between Savers and Borrowers: Banks as middlemen between savers and borrowers; "Nonbanking" middlemen between savers and borrowers. Creation of Credit by Commercial Banks: Potential creation of credit by an individual bank; Limits to credit creation by an individual bank; Credit creation by the banking system; Contraction of credit by commercial banks.
40. THE PRICE LEVEL; OR THE VALUE OF MONEY 600
What Is Meant by the "Value of Money"? Price Indexes and Price Level Changes. Price Level Changes and the Distribution of Income: The position of creditors; The position of debtors.
41. SPENDING, TRADE, AND THE PRICE LEVEL 612
The Equation of Exchange: Statement of the equation; Interrelations of the factors in the equation; Restatement of the equation. The Quantity of Money (M): Determinants of the quantity of money; The quantity of money and inflation. The Quantity of Credit (M'). The Velocity of Circulation of Money and Credit (V and V'). The Volume of Transactions (T). Interdependence of Spending, Trade, and the Price Level.
42. GOLD AND THE PRICE LEVEL 627
The Gold Standard. The Abandonment of the Gold Standard. "Too Little" and "Too Much" Gold: The problem of "too little" gold; The problem of "too much" gold. Gold, Silver, and the "Mixed Standard." Future Policy in Regard to Gold.
43. THE FEDERAL RESERVE SYSTEM 636
Organization of the Federal Reserve System. Service Functions of the Federal Reserve: Holding member bank reserves; Furnishing money for circulation; Clearing and collection of checks; Supervising member banks; Fiscal agent for the federal government. Credit Control by the Federal Reserve: Open market operations; Rediscount rate charges; "Advances"; Direct pressure; Changes in member bank reserve requirements; Changes in "margin" requirements; Control of installment credit; Credit policy in wartime; Summary.

PART IX

GENERAL BUSINESS FLUCTUATIONS AND ECONOMIC POLICY

44. GENERAL BUSINESS FLUCTUATIONS 657
What Are "Business Cycles"? Entrepreneurs and the Level of Business Activity. The Framework of the Cycle. Depression. Revival. Prosperity. Recession. Conclusion.
45. ECONOMIC POLICY AND BUSINESS FLUCTUATIONS 682
The Basis for Economic Policies to Mitigate Business Fluctuations. Monetary Policy: Issue and withdrawal of money; Treasury control of bank credit; Federal Reserve control of bank credit; The problem of "timing" monetary policy; Federal Deposit Insurance Corporation.

45. ECONOMIC POLICY AND BUSINESS FLUCTUATIONS (*Cont.*):

Fiscal Policy (Government Finance): Fiscal powers to check inflation; Fiscal powers to stimulate revival; Timing of fiscal policy; Fiscal policy and the government debt. Policy Toward Costs. Miscellaneous Policies. "Long-Run" Stabilization. "Business Cycle" Policy in War-time.

PART X

THE PUBLIC ECONOMY

46. INTRODUCTION TO THE PUBLIC ECONOMY 707

Government Expenditures and Revenues: Expenditures and revenues in recent years; The historical trend of expenditures and revenues. Social Policy and the Public Economy: The "ends" of social policy in the public economy; Special characteristics of the public economy; Consumers' sovereignty and citizens' sovereignty.

47. PUBLIC EXPENDITURES 719

The Scope of Government Expenditures: The principle of maximizing "social welfare"; Money and real transfers in the public economy; The government as transfer agent and as entrepreneur. The Distribution of Government Expenditures: "Regular" expenditures; War expenditures; The distribution of expenditures by income groups.

48. PUBLIC REVENUES: TAXATION 733

"Good" and "Bad" Taxes: Social criteria for taxes; "Equity" in taxation. Distribution of the Tax Burden. Shifting and Incidence of Taxation. Excises: Incidence; Evaluation; Motor fuel taxes. Retail Sales Taxes: Incidence; Evaluation. Pay-Roll Taxes: Incidence; Evaluation.

49. TAXATION (CONTINUED) 760

Personal Income Taxes: Incidence; Evaluation. Death and Gift Taxes: Incidence; Evaluation. Corporation Income Taxes: Incidence; Evaluation. The Taxation of Property: Administration; Taxes on personal property; Taxes on "improvements on real property"; Taxes on real property ("land"); Severance taxes; Capitalization; Evaluation of property taxes. Special Assessments. Conclusion.

50. FISCAL POLICY AND THE NATIONAL INCOME 790

Revenues: Taxation; Borrowing; New money; Allocations, individual price-fixing and rationing. Expenditures: Expenditures in periods of relatively full employment; Expenditures in periods of widespread unemployment. Integration of Federal-State-Local Fiscal Activities.

51. SOCIAL SECURITY: A CASE STUDY 809

The Background of Social Security. The Federal Social Security Act. Unemployment Insurance: The legal status of unemployment insurance; Is "unemployment insurance" really insurance? Economic aspects of unemployment insurance. Old-Age Insurance and Aid to Special Needy Groups: Old-age pensions; Aid to other special groups. Workmen's Compensation. Conclusion.

PART XI

THE INTERNATIONAL ECONOMY

CHAPTER	PAGE
52. INTERNATIONAL TRADE AND LENDING	829
The Basic Case for International Trade: Interregional and international trade; The law of comparative advantage; The law of comparative costs, free trade, and the price system. The Basic Case for International Lending. International Community and Conflict of Interests. Political Aspects of International Trade. The Balance of Payments.	
53. FINANCING INTERNATIONAL TRANSACTIONS	845
Foreign Exchange. Exchange Rates: Exchange rates with "free" currencies; Exchange rates under the gold standard; "Controlled" exchange rates. International Adjustments: International adjustments under the international gold standard; International adjustments under fluctuating exchanges.	
54. INTERFERENCES WITH INTERNATIONAL TRADE: THE TARIFF	856
The Background of the Tariff. Development of the Tariff. Effects of the Tariff: The "infant industry" argument; The "national self-sufficiency" argument; The "favorable balance of trade" argument; The "protect home industry" argument; The "equalize costs of production" argument; The "protect our high wages and high plane of living" argument; The "increased employment" argument; The "national economic stability" argument; Conclusion. Tariff Making in the United States.	
55. INTERFERENCES WITH INTERNATIONAL TRADE: QUOTAS, EXPORT SUBSIDIES, AND EXCHANGE CONTROL	874
Quotas and Allied Devices. Export Subsidies. Exchange Control. Conclusion.	

PART XII

AGRICULTURE: A CASE STUDY

56. AGRICULTURE: A CASE STUDY	883
The Farmer's Economic Status in Society. Changing Fortunes in Agriculture: Agriculture up to the 1920's; Agriculture during the 'twenties. Reasons for the Slow Downward Adjustment of Agricultural Production to Declining Demand: Real estate and associated fixed costs; Farm labor and associated fixed costs; Production curtailment and inelastic farm demand. Agricultural and General Business Fluctuations. Agriculture and International Trade. Agriculture's Claim to Special Aid. Farm Aid—1921-1933. Farm Aid Programs Since 1933: Production control and marketing quotas under AAA; Crop loans and storage; The surplus disposal program; The trade agreements program and export subsidies; The farm credit program; Additional direct aid to farmers; Efforts to promote more efficient use of resources. Conclusion.	
INDEX	917

Part I

LAYING THE FOUNDATIONS

CHAPTER 1

The Problem of Economizing

PEOPLE everywhere strive to obtain what they anticipate will bring them satisfaction; but on whatever part of the globe they may be and at whatever time, most of the means for satisfying these wants are available only in limited amounts, in quantities less than the people desire. Indeed, few things in our lives today are available in such quantities that everyone can have all he wants of them free. Even fresh air is something for which people pay in urban rents.

A problem of "economizing" arises when people want more of a thing than is freely available, and when there are alternative uses to which it may be put or different rates at which it may be consumed over a period of time. Individuals face a problem of economizing in the use of scarce time, money, and energy. If we spend our time in one way we must forego spending it in another; for most of us time is scarce and there are many alternative uses to which it may be put. Businesses face a problem of economizing in the use of machinery, labor, and raw materials. Society as a whole faces the problem of determining how the many resources of the whole society will be used.

Most of the goods which we as individuals want to consume (shoes, coats, houses, automobiles, taxis, movies, food) are made with the services of land, machines, raw materials, human labor—resources that are limited in quantity and that can be used in a number of different ways. Both the goods and the resources that contribute to producing them are "economic" in that there is a problem of economizing in their use. Moreover, all these economic goods and services are also transferable; they *can* be bought and sold, though they are not necessarily thus transferred from one potential user to another.

Many attempts have been made to define precisely the field of economic study. It is always difficult to frame definitions to which no exceptions can be taken. Here we are interested only in setting forth the core of economics.

Three conditions characterize the goods and services the economizing of which is the central problem in our study of economics: (1) scarcity, (2) alternative uses, and (3) transferability. Something is "scarce" when people want more of it or of the things it can help produce than is freely available. Thus even in years of large crops when prices are

low, wheat will be "scarce" in that there will not be as much of it as people would like to have if they could get it for nothing. But scarcity alone is not enough to create a problem of economizing; there must also be some choice with regard to how a consumer's good or a productive resource will be used. The choice may be one of different uses at a given time, or one of spreading the use in different ways over a period of time. Finally, in order to be strictly economic, goods and services must be transferable from one individual to another. Through transfers there is the possibility of exchanging the goods and services either for money or for other goods and services. The emphasis placed on transferability is of great importance because we are interested primarily in social rather than individual economy.

The functioning of a society in its use of the total economic resources available to it is the core of economic study, and many broad problems are associated with this core. In the following sections of this chapter we shall briefly examine first, the selection and consumption of goods in a society; second, the nature of the resources with which we work; third, the role of economic organization in bringing together scarce resources for the production of the desired goods and services; and fourth, some of the interests of individuals in the functioning of the economic system.

Production for Consumption

You and I and all the rest of the people are the ultimate consumers of the goods and services produced by the total economy. We buy and use food, clothes, houses, movies, books, and many other things. It may appear at first sight that we work to get "money"; but why do we want money? Unless, like King Midas, we want primarily the satisfaction of counting our coins, we want the money for what it will buy. Thus in our capacities as producers we seek money to spend on goods and services that we should like to consume. How do we get the money? Each of us gets it by producing things other consumers want to buy. The entire economy is thus ultimately directed toward providing the things we want as consumers. It may operate well or poorly to this end; nevertheless, our motivations in economic life may be largely traced back to our desires for satisfactions derived from consumption of goods and services.

Scales of preferences, planes of living, standards of living

What do we want and what do we get?

You might make out a list of the economic goods and services you want, ranking them in the order of their importance. This would become very complex if it were at all complete. There would be a lot of alternatives. Perhaps you would prefer a \$1,000 car to a \$1,000 vacation trip; but you might prefer a \$700 car and a \$300 trip to either a \$1,000 car or a \$1,000 trip alone. The list of your many preferences arranged in order of importance is called a "scale of preferences." It is with such a scale vaguely in mind that you make your actual selections of goods and services.

The economic goods and services you actually consume constitute your

individual "plane of living." The total of economic goods and services actually turned out and used by a whole economy constitutes the plane of living of all the consumers in that economy viewed collectively. The plane of living is thus a composite of all the economic goods and services, big or small, important or unimportant, which are actually consumed.

Scales of preferences and realized planes of living are to some extent related: On the one hand, scales of preferences affect planes of living, while on the other, planes of living influence scales of preferences. Those persons who are in a position to influence the ways in which resources are used will exert their influence in accordance with their preferences. Such volition may be exercised by individual consumers in America; whereas in totalitarian countries a dictator may determine the consuming habits of his subjects. In either case preferences will affect the ways in which resources are used and will influence the composition of the planes of living which result. The influence of planes of living on scales of preferences is equally obvious. Your individual preferences will depend in part on your past experiences with items composing your own planes of living, and in part on your knowledge about the items in the planes of living of other people. Changes in the composition of these planes will often lead to changed ideas of the relative importance of various goods and services; in short, to changes in scales of preferences.

But in stating that planes of living influence scales of preferences we have by no means explained the composition of scales of preferences, nor can we fully explain them here. For we must look back of particular preferences to the general satisfactions we expect to derive from our consumption of goods and services. We want physical comfort in living—in eating, sleeping, bathing, visiting, reading, traveling. We want aesthetic surroundings, excitement and entertainment, security and peace, intellectual stimulation, friendship and affection, and many other highly subjective satisfactions. Economic goods and services sought in accordance with scales of preferences are means contributing to some of these satisfactions. Some of these satisfactions, on the other hand, can be realized only through noneconomic channels. Yet the sum total of the satisfactions sought may be described as the "standard of living" of an individual or of a group. Hence, in summary, we may conclude:

1. A standard of living is the sum of the satisfactions sought by an individual or group. Economic activity can contribute to the realization of some of these satisfactions.

2. A scale of preferences is a ranking of specific goods and services in the order of their importance to an individual or a group.

3. A plane of living is the sum total of the goods and services an individual or a group actually consumes.

Thus it is the standard of living that leads us to have certain preferences for particular kinds of economic goods and services, which we then attempt to obtain in one way or another by economic activity. The resulting productive activity turns out goods and services that make up

the actual plane of living. Lastly our experience with that plane and with its capacity to yield satisfactions in turn influences both our scales of preferences and our standards of living.

The meaning of the phrases "standard of living," "scale of preferences," and "plane of living" may be illustrated by some attitudes toward one phase of a standard of living: the satisfactions to be obtained from bathing, and the corresponding rank of bathtubs in scales of preferences.

In a book called *Are We Civilized?* Robert Lowie describes the attitudes of various peoples toward cleanliness.¹ He says of the Eskimos,

As rats leave a sinking ship, so vermin will leave a dying man according to Greenland theory. Hence a louseless Eskimo is ill at ease.

However, of another primitive tribe he reports:

Day in and day out the Cheyenne of our Western Plains took their plunge, even when they had to break the ice in the wintertime.

Baths were suspect in Europe in the Middle Ages. Of this civilization Lowie says,

To be sure, Christianity did not oppose baths per se—so long as they were intended to serve cleanliness of health and not sheer pleasure. But the ascetically-minded naturally set up lofty standards. One saint was for letting the sick bathe as often as they pleased, but others—especially the young—were to do so rarely, and monks were to be allowed ablutions at Christmas and at Easter.

Yet public baths became popular because,

. . . the crusaders got into touch with Oriental customs and were stimulated into more frequent indulgence. Hence sprang the great public establishments of the Middle Ages which became truly popular and indispensable institutions. . . . As in ancient Rome, so in medieval Europe the public baths became an institution with very wide functions indeed. . . . Why bathe at home when one could so much more agreeably meet pleasant company at the sweat-house?

Apparently the satisfactions to be derived from bathing vary according to customs and conditions. Standards of living, insofar as they relate to bathing at least, show wide differences among these different groups.

Even within the past century in America there have been important shifts in the attitude toward bathing and the satisfactions associated with it. In an old physiology text published in 1848 appear the following remarks:

We bathe the face and hands daily, and oftener, and know how comfortable the skin upon these parts feels after this operation. But if this duty is neglected, the skin is irritable and irritated; it seems stiff and loaded, and we feel disposed to scratch and rub it to remove the disagreeable burden. But the other parts, which are not so frequently washed, are not so easily offended. They bear the burden of accumulated excretions and dust with less complaint. But if they were cleansed as faithfully as the hands and the face, they would be equally sensitive, and feel as keenly the comfort of a bath and the discomfort of neglect. This sensibility of the skin of the hands and face is a mere matter of cultivation,

¹ Lowie, Robert H., *Are We Civilized?* Harcourt Brace & Co., New York, 1929, p. 235 ff.

and might as well be cultivated in the skin of the other parts which are covered with clothing. But those parts which are not exposed to sight are with most people rarely, and with some never, bathed; and the great majority of mankind leave so much of their surface untouched with water, from summer, through the entire cold season, until summer again returns. . . ."

It is desirable that every one should be able to take, daily, a plunge bath. But this is impossible for all. The convenience of a large bathing tub cannot be provided in every house. Yet a good substitute is within the reach of all. The sponge bath is very easily taken, and requires but a very limited and simple apparatus.²

These last two paragraphs reveal something about the *standards of living* and the *scales of preference* and *planes of living* in this country in 1848 insofar as they relate to bathing. The purpose of the writer quoted was to urge a standard of cleanliness "higher" than was then common. In the first paragraph, he points out the satisfactions and benefits of frequent bathing, and laments the "low" standard of living represented by the annual bath. In both paragraphs he implies the low position of bathtubs in scales of preferences. The second paragraph reveals that the bathtub was not a common item in the planes of living of that day. This is confirmed by Godey's *Lady's Book* for the same period. The issues for 1848-1850 contain many house plans, but in only one is a bath mentioned:

In the floor of the dining room there is a bath, the lid of which opens like a trap-door, and the descent is by steps. The bath is supplied with hot water from the boiler in the scullery, and with cold water from a cistern in the back-kitchen.³

Here then is an example of one particular phase of consumption. We might instead have selected the heating of homes, transportation and communication, dramatics, music, or art. Whatever the particular examples selected, we would have found that there have been shifts over time, and that there are differences between individuals and between groups of people in these respects today.

Present American planes of living

Planes of living, since they can be described and measured in terms of goods and services consumed, are easier to observe than are scales of preferences or standards of living. The composition of planes of living varies greatly as between different civilizations and times. Even in a given civilization at a given time the planes of different individuals and groups vary with differences in size of income, in degree of security of income, in occupation, in social and cultural inheritance, in climatic conditions, in family composition, in personal needs and tastes. Government surveys have made available considerable information relating to items composing planes of living in the United States today.

The data given in Table 1—1 show the average amounts spent per

² Jarvis, E., *Practical Physiology for the Use of Schools and Families*, Thomas Cowper, Thwait publishers, Philadelphia, 1848, pp. 230-231, 233.

³ Godey's *Lady's Book*, Philadelphia, May 1850, "Model Cottage," page 353.

TABLE 1-1 *

AVERAGE FAMILY EXPENDITURES FOR CONSUMPTION AT DIFFERENT INCOME LEVELS, 1935-36

	<i>Food</i>	<i>Housing</i>	<i>Household operation</i>	<i>Clothing</i>	<i>Auto</i>	<i>Medical and personal care</i>	<i>Recreation and tobacco</i>	<i>Furnishings and equipment</i>	<i>Reading and education</i>	<i>Other</i>
Under \$500	\$ 203	\$ 90	\$ 57	\$ 35	\$ 15	\$ 31	\$ 15	\$ 9	\$ 6	\$ 5
\$ 500- 750	310	125	85	56	28	43	25	16	9	10
750- 1,000	380	161	106	78	44	56	36	27	13	13
1,000- 1,250	433	203	130	100	70	71	47	38	18	17
1,250- 1,500	487	230	149	123	93	84	58	48	23	21
1,500- 1,750	527	267	166	147	123	103	71	56	26	26
1,750- 2,000	558	302	186	164	154	114	82	68	31	25
2,000- 2,000	617	349	213	207	200	133	100	76	40	33
2,500- 3,000	680	404	260	255	242	158	122	84	52	35
3,000- 4,000	770	485	319	316	289	186	153	102	64	45
4,000- 5,000	852	571	400	408	382	224	189	110	88	52
5,000-10,000	1038	784	584	557	522	337	268	158	124	82
10,000-15,000	1214	1204	761	829	681	341	419	227	284	137
15,000-20,000	1785	1490	1179	1265	919	572	590	277	606	451
20,000 and over	2261	2721	2177	2177	1759	1088	1047	461	628	503

* Adapted from: National Resources Committee, *Consumer Expenditures in the United States, 1939*, Table 2, p. 23.

TABLE 1-2*

PERCENTAGE OF TOTAL EXPENDITURES OF AMERICAN FAMILIES FOR MAIN CATEGORIES OF CONSUMPTION BY INCOME LEVEL, 1935-36

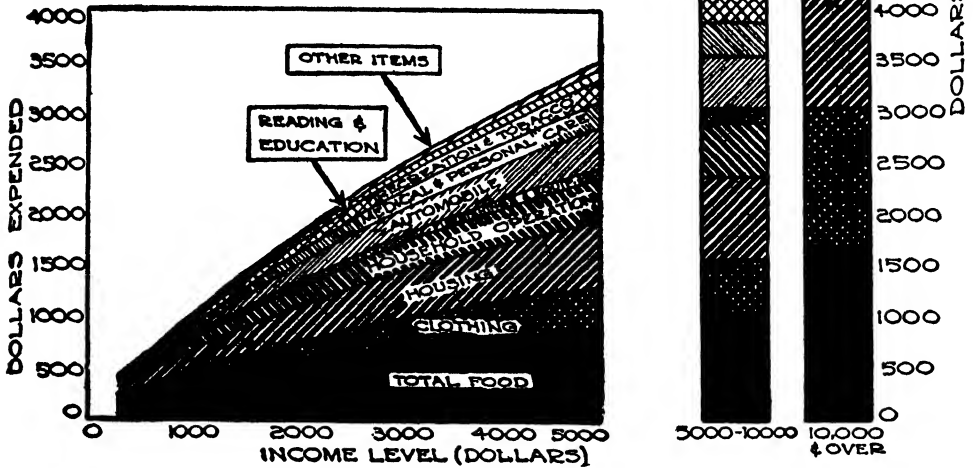
	<i>Food</i>	<i>Housing</i>	<i>Household operation</i>	<i>Clothing</i>	<i>Auto</i>	<i>Medical and personal care</i>	<i>Recreation and tobacco</i>	<i>Furnishings and equipment</i>	<i>Reading and education</i>	<i>Other</i>
Under \$ 500	43.6	19.3	12.3	7.5	3.2	6.6	3.2	1.9	1.3	1.1
\$ 500-750	43.8	17.7	12.0	7.9	4.0	6.1	3.6	2.3	1.2	1.4
750-1,000	41.5	17.6	11.6	8.5	4.8	6.2	4.0	3.0	1.4	1.4
1,000-1,250	38.4	18.0	11.5	8.9	6.2	6.3	4.2	3.4	1.6	1.5
1,250-1,500	36.9	17.5	11.3	9.3	7.1	6.4	4.5	3.6	1.8	1.6
1,500-1,750	34.9	17.6	11.0	9.7	8.1	6.8	4.7	3.7	1.7	1.8
1,750-2,000	33.1	17.9	11.0	9.7	9.2	6.8	4.9	4.0	1.9	1.5
2,000-2,500	31.4	17.7	10.8	10.5	10.2	6.7	5.1	3.9	2.0	1.7
2,500-3,000	30.0	17.6	11.3	11.1	10.5	6.8	5.3	3.6	2.3	1.5
3,000-4,000	28.2	17.8	11.7	11.6	10.6	6.8	5.6	3.7	2.4	1.6
4,000-5,000	26.0	17.4	12.2	12.5	11.7	6.8	5.8	3.4	2.6	1.6
5,000-10,000	23.3	17.6	13.1	12.5	11.7	7.6	6.0	3.5	2.8	1.9
10,000-15,000	19.9	19.7	12.5	13.6	11.2	5.6	6.9	3.7	4.6	2.3
15,000-20,000	19.6	16.3	12.9	13.8	10.1	6.3	6.4	3.0	6.6	5.0
20,000 and over	15.3	18.4	14.7	14.7	11.9	7.3	7.0	3.1	4.2	3.4

* Adapted from: National Resources Committee, *Consumer Expenditures in the United States, 1939*, Table 7 A, p. 78.

family at different income levels on various groups of commodities. These data are illustrated graphically in Figure 1—1. They may be compared with the figures indicating percentages of family expenditures for different commodity categories, as given in Table 1—2 and Figure 1—2. All items increase in absolute amounts with higher incomes, but some increase much more than others. Though they increase in absolute amount, the expenditures on food absorb a much smaller proportion of large than of small incomes. Automobiles, on the other hand, take a much larger per cent of high incomes.

These data are indicative of the way in which the budgets of families in different income groups are divided among major consumption categories. To interpret the data as

Fig. 1—1. Average expenditures for consumption at different income levels, 1935-1936.



indicators of the composition of American planes of living we may take one more step. In what income classes do most American families fall? In the year 1935-36, approximately half the families in the United States had incomes under \$1,070. If the population is divided into three groups, the income estimates are roughly those given in Table 1—3.

TABLE 1—3*

FAMILY INCOME DISTRIBUTION IN
THE U. S., 1935-36

Lower third	Under \$780
Middle third	\$780-1450
Upper third	Over \$1450

* Source: National Resources Committee, *Consumer Incomes in the United States*, 1939, p. 9.

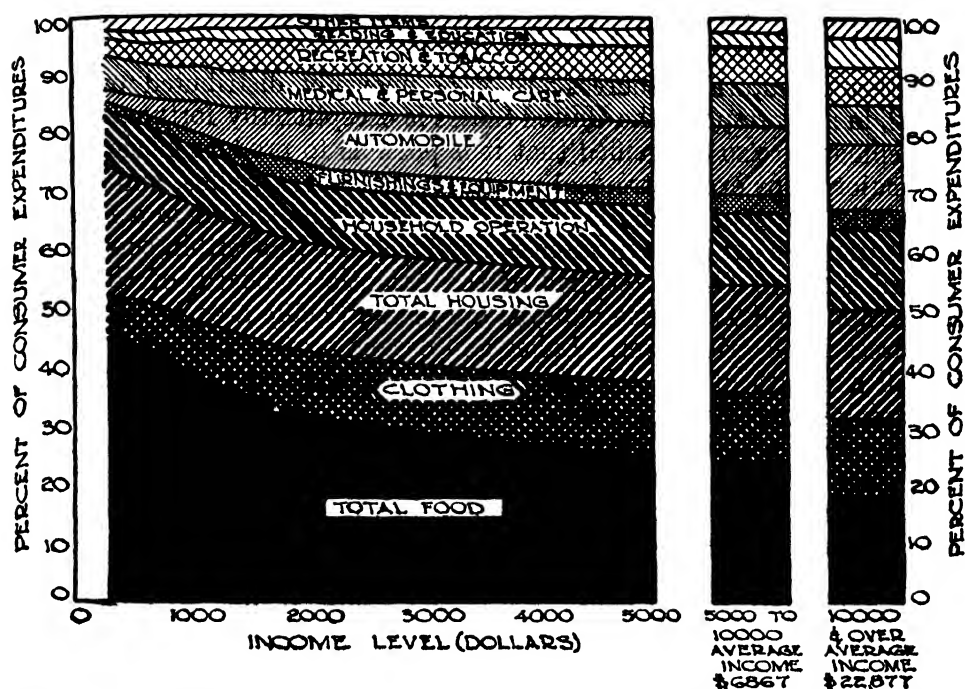


Fig. 1—2. Proportionate distribution of average expenditures for consumption at different income levels, 1935-1936.

Economic Resources and Technology

The central economic problem of any society is the problem of "economizing" scarce resources in order to produce the various goods ultimately desired by consumers.

Concretely, the economic resources of a society take thousands of forms. Natural resources (fertile agricultural lands, forests, mineral deposits, waterpower) basically condition and limit the productive possibilities of the total economy. Natural resources, however, are only a part of what is needed for production. There must be labor power to exploit these resources, and there must be equipment produced, which may be used for further production. The efficiency of the equipment will depend not only on the natural and human resources but on the state of technology. Thus the wealth of a society will depend on all its resources and on the efficiency with which they are combined in production both technically and through the society's organizing processes.

The resources on which the members of a particular nation may draw are worldwide; but barriers to their use may be imposed by the various nations. Any attempt to note concretely what these resources are would be an abbreviated almanac of world resources. We shall, therefore, take a more limited view and examine the resources of the United States, recalling, however, that resources outside the United States are drawn upon by this country just as American resources are drawn upon by other countries.

Natural resources of the United States

Although natural resources are very unevenly distributed over the world, in most respects the United States is exceptionally fortunate. Soil and climate are so diversified that it is possible to produce efficiently all the major crops except tropical products such as tea, coffee, and rubber. Coal, oil, iron, copper, and many of the minerals important in a machine civilization are available in extensive and rich deposits within continental United States. The waterways of the Atlantic seaboard, of the Ohio, Mississippi, and Tennessee valleys, and of the Pacific area provide great resources in electric power. Lakes and oceans provide sea food. There is still a wealth of timberlands although conservation of these areas is becoming increasingly important.

The deficiencies in our natural resources are revealed by our imports. These are mainly of tropical products and sugar. Sugar could be grown here; in fact we do produce beet sugar, but it is much cheaper to buy sugar from Cuba and other semitropical areas. With the exception of rubber, most of the tropical products might be regarded as relatively nonessential in the total of our economic activity, though we might miss products such as coffee very much indeed. The inadequacy of resources to produce these goods may therefore be regarded as of only limited importance in our economic life. Some minerals which occupy key positions in modern industry (especially for the production of high-grade steel used in machine tools) are either wholly imported or are imported in very large proportions. This dependency is exaggerated by the needs of wartime. The most important examples are nickel, tin, and manganese.

The natural resources in the United States have increased in effectiveness with the opening up of transportation facilities and with research into uses of new materials and improved and new uses of old ones. At the same time, evolving techniques of production call for an ever-increasing range of materials, some of which are not available within our national boundaries. Thus simultaneously we grow richer in effective resources and more dependent on foreign sources of supply of a few particular products. The wealth of the country and of the world depends, therefore, not only on what lies in hills, and plains, and streams, and oceans; it depends also on which of the natural resources we know how to use and how well we know how to use them.

Produced resources of the United States

The United States is rich not only in natural resources but also in equipment with which to carry on production. The National Resources Committee estimated the total value of productive equipment in 1935 in the United States to be about 190 billion dollars, and the value has since increased. If this equipment were destroyed it could be reproduced only at very great cost and over a period of many years. Much of this equipment is highly specialized in the uses to which it may be put: watch-making machines, shoe machinery, houses, dairy plants—equipment of a

million kinds. Extensive changes in production can be made only slowly because special equipment is needed and would have to be produced. Existing forms of equipment are frequently not adaptable to the new uses. Without its existing railroads, factories, houses, farm buildings, and so on, the United States would for many years and decades to come be quite incapable of providing the plane of living known to the present generation.

Manpower in the United States

Without men to operate them, potential farm lands, mines, or machines will produce nothing. It is all these resources *in combination* that produce the goods and services that constitute the planes of living of the population. We might measure the extent of the nation's manpower by taking census figures, which show that the total number "in the labor force" in the United States at the time of the 1940 census was 52,840,762. This includes both employed and unemployed workers (the active and the potential workers) in "gainful" employments. In addition there were 28,838,484 reported as doing housework in their own homes and not engaged in any gainful employment. But the number of men or manhours available is by itself a very poor indicator of the nation's richness in manpower since it ignores the important matter of abilities and training. The real productivity of the manpower of the nation depends on brains and brawn, on people trained in a million different areas of activity and equipped with all degrees of skill and efficiency.

Technology

Our civilization is frequently described as a "machine civilization," a term which connotes the technical inventiveness that has been so outstanding. The invention of ever more effective production techniques and instruments explains to a large degree the planes of living attained in recent generations. What can be produced with given resources depends on our knowledge of the technological possibilities in the use of the resources. The available resources themselves are determined by such technological knowledge; the possibility of utilizing at all many things depends on such knowledge; and the knowledge determines in part the efficiency with which any existing resources can be used. For example, many of the minerals now regarded as the most prized possessions among our resources were unknown or had little importance a century or two ago. The most skilled spinners on the old-fashioned spinning wheel were far less productive than are modern unskilled workers combining their efforts with machine equipment made possible by our greater technical knowledge. The advanced application of technology to production and its effects on the resulting planes of living of the people are two of the **most** striking characteristics of modern Western civilization.

The Economic System

To recognize that on the one hand we have scales of preferences **as to** what we want to realize in our planes of living and that on the other hand we have scarce resources and some degree of knowledge of various tech-

nologies for combining these resources productively is not enough; these economic resources are not used in a social vacuum. In every society there is some set of social processes for the selection of channels into which limited resources will be directed and for the determination of the techniques for combining them. All social processes (family living, religion, recreational customs, and so on) have at least an indirect effect on the processes of combining resources to produce a plane of living; but some processes are more directly focused on the performing of this function than are others. In our society these direct processes include, for example, the buying and selling of raw materials, labor, consumers' goods, and other goods and services; they include "business activities" of almost all kinds. The social processes most directly determining the ways in which resources will be combined are the "economic aspects" of social structure. These aspects are sometimes roughly designated as the "economic system"; they constitute the structure through which the society as a whole "economizes" in the use of its resources. An oversimplified statement of the relation of this "economic system" to the elements it coordinates may be presented diagrammatically as follows:

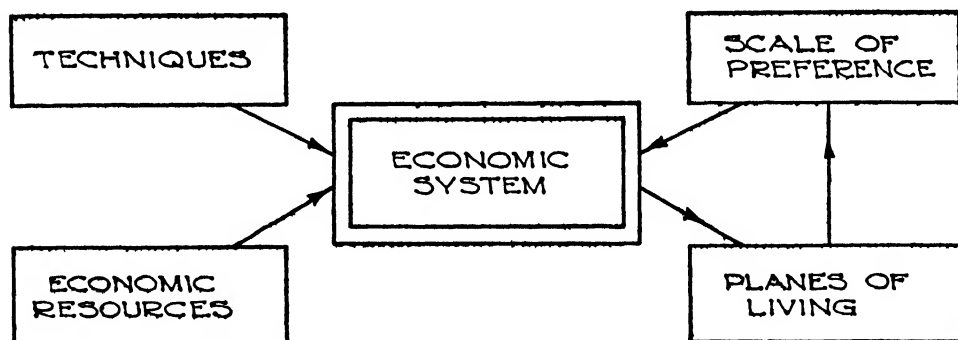


Fig. 1—3. The coordinating function of the economic system.

Under the impetus to provide the desired goods and services, decisions are made as to how scarce resources will be used, and what techniques will be employed. The particular ways in which these decisions are made may differ somewhat in different societies. It may be that a central authority dictates who shall do what, and how goods will be divided. Or there may be a system of pricing in a free market, such as is described roughly in Chapter 2. But in all societies there must be some method of determining how resources will be allocated among various uses. Whatever the method may be, this task is the central function of an economic system.

Place of the Economic System in the Achievement of a Standard of Living

Standards of living have been defined here as constituting a total of satisfactions that are considered important. Some of these satisfactions

may be obtained by the use of the economic goods and services which constitute a plane of living. But it is only insofar as a plane of living can contribute to the realizing of a standard of living that economic activity contributes to the achievement of that standard.

Some satisfactions, such as those derived from attaining prestige in the eyes of the community, are realized through both economic and non-economic means. Living in a fine house will have its effect; so will belonging to the "right" club, and being born in the "right" class, of parents of the "right" race and religious creed. The social system provides a wide range of channels both economic and noneconomic for the attainment of prestige.

Other satisfactions may be achieved almost entirely through non-economic channels. Among the satisfactions sought will be, for example, psychological security in a close personal relationship with a friend. Obviously no plane of living and no part of a plane of living can fully provide this satisfaction. It depends on the availability of someone to fill this need, and when that person is found the function that he performs cannot be performed in the same way by anyone else, nor can he provide this service to just anyone, even though the latter might be willing to pay for it. His services as a friend are noneconomic. Associations of people in formal or informal groups provide satisfactions of friendship and sociability which cannot be provided merely by a large income. Associations may be between two individuals. Perhaps a young man has a very good voice for the use of which Metropolitan is willing to pay him large amounts; if so his ability is an economic resource. But suppose he uses this voice to serenade his sweetheart. She enjoys it much more because his voice is beautiful than were he to croak his love for her, but no other voice, however beautiful, could be a substitute for his. The personal aspect of his serenade removes it from the plane of living and it becomes instead a part of the other means by which standards of living may be realized.

The social system which provides these various channels for the realization of a standard of living also conditions the ways in which economic activities themselves will be carried on. Procedures developed in one sphere of action are tied up with activities in other spheres. Examples are limitless. Thus we may set up rules of "honesty" which we assume will be followed in the various ranges of human action, and in some cases we introduce coercive means of enforcing "honest" behavior. Certain days are national or religious holidays for us, and only certain kinds of work will be done on these days. Because of the abolition of slavery the old plantation system of the South has been disappearing, thus changing the form of economic activity and social life. And so we might continue.

The particular forms which economic activity will take are influenced by the other aspects of social activity and in turn play some part in influencing the forms which that social activity will take. There is, in short, a mutual interdependence between the economic and the other

parts of the total social structure. The "economic system" is but one of the systems of social organization directed ultimately to the fulfillment of standards of living.

Individuals and the Economic System

In the functioning of an economic system individuals play various roles. They are consumers of its products and they contribute to the creating of those products. They are also receivers of incomes, which they expend and thereby make their claims to the products of the economy. How do individuals view the economic system from each of these roles? Some examples may prove revealing:

1. The man controlling the policies of a manufacturing enterprise is interested in how the system functions in its effects on the price of his product, on the wages he will pay his employees, on the cost of his equipment. He is interested in possible competition from producers of substitutes, including those in foreign countries. He is interested in the effects of government action designed to prevent him and others from forming combinations to keep prices up. He is interested in the effects of trade-union activities in his own concern and in the plants of rivals. He is concerned with interest rates and the markets for stocks and bonds. A host of economic problems are his daily concern.

2. The farmer has interests in the economic system which are in some respects similar to those of the manufacturer. He is interested in factors determining the prices of what he has to sell and of what he must buy. He is especially concerned with the effects of various economic changes on the value of land. He is interested in problems related to the financing of farm mortgages. Developments in world trade and their implications for the sale of farm products are of great concern to him.

3. The investor, whether in corporate securities, government bonds, or some other form of property, is interested in changing price levels, interest rates, the fortunes of particular enterprises and industries. All of the economic forces affecting returns on his investments become matters of special concern to him.

4. The laborer has obviously vital interests in the position of laborers as income receivers—in the growth and policies of trade-unions, the relation between wages and retail prices, the policies of government and their effects on employment, the displacement of workers by machines, the effects of immigration, or of tariffs, on the position of labor. These matters are of particular importance to labor leaders in evaluating various policies that they may pursue, whether in dealing with employers or politicians. Labor leadership requires a comprehensive understanding of economic processes.

5. The interest of the individual as a consumer is common to all of us, but in each particular case it is likely to appear far less important than his interest as an income receiver. As consumers we are most immediately interested in practices which mislead us and which make prices of

particular goods high—for example, in misleading advertising, in costly merchandizing, in the effects of tariffs and sales taxes. Our interests as consumers spread over a whole range of problems each of which is the especial interest of one or more particular producer groups.

Economic Analysis in the Analysis of Social Organization

The various broad fields of study and research which have come to be classified as natural sciences, social sciences, and philosophy, all contribute to an understanding of the activities we have just described. Some fields of study contribute to an understanding of scales of preferences: what they are, how they are determined, and the effects of various standards of living and planes of living on personal development and living. The contribution of philosophy, and in particular of that branch of philosophy known as ethics, is mainly in the examination and analysis of the goals and desires which make up the standard of living. Philosophy also analyzes and interprets scales of preferences and planes of living with a view to determining what the contents of these planes contribute to satisfactions. Psychology, physiology, and other sciences contribute to our understanding of what lies back of a standard of living and of how planes and standards are related. Application of the contributions of these fields is made through home economics and other technologies. Thus, dietetics draws on chemistry and physiology to show what elements in a plane of living would effectively satisfy some of the basic desires in the standard of living of the individuals involved.

A study of productive resources and technology is no less important. The nature of available resources and technical possibilities in their uses receive the special attention of technologies, such as engineering and certain aspects of home economics. These again are based on pure sciences such as chemistry, physics, or biology. Thus the electrical engineer derives his analytical techniques and practical judgments largely from physics; the home economist applies laws of physics and chemistry to cooking or textile manufacture.

Procedures by which human activities in the use of resources are organized in response to our preferences constitute the special field of study of the social sciences: economics, sociology, political science, and certain aspects of psychology. Political science examines the role of the state in the control of social action. Sociology is primarily concerned with various aspects of human interrelation such as the operation of customs and mores in a society, or the functioning of groups. Economics examines another aspect of social organization—the system by which limited resources are allocated to the attaining of many and diverse ends expressed in scales of preferences.

The special contributions of these different social sciences may be illustrated by the use of an example. How, for instance, would specialists in each of these fields attempt to interpret trade-union activity? The political scientist studies trade-unions as a power group functioning in a particular governmental setting. He realizes that they may be provided

for directly within the legal structure or that they may exert an outside pressure on that structure and on the division of powers in the state, since they are, among other things, channels for the expression of public opinion of certain groups. The sociologist is interested in the functioning of trade-unions as weapons in the struggle between social classes, as devices for increasing the prestige of workers in a certain society, and as organizations restricting through group pressure the opportunities for advancement of particular laborers.

The economist is interested in trade-unions because of the way in which they influence what is produced, how much, and by whom. He is interested, also, in how trade-unions affect people's incomes, for he knows that people with higher incomes have more dollars to spend and will, therefore, get more of the total of goods and services. Moreover, to the economist the wages of workers are costs of production, and he knows that the way in which a society solves many of the problems associated with organizing the use of economic resources to produce the goods and services that make up planes of living will depend among other things on the activities of trade-unions in that society.

CHAPTER 2

The Modern Exchange Economy

MAN lives by cooperating with his fellowmen. There is no truth in the entire economic realm which is of more basic importance than this. In our modern exchange economy this cooperation is indescribably broad and complex; even a brief look around will reveal many of its ramifications. We depend for the satisfaction of our simplest wants upon the efforts of innumerable people in all parts of the world, as well as upon the fruits of the labors of past generations of men and the material resources with which they worked. Yet this vast cooperation has been essentially undesigned and uncontrolled by man; there has been no human authority presiding over this total complex of activities, organizing them all and directing them to the common ends which they serve.

Man can organize, and indeed he has organized much. Tens of thousands of workers are employed in some of our large industrial plants, and often many of these huge plants are joined together in a single organization. The American Telephone and Telegraph Company, for example, spreads over the entire United States, with some 275,000 employees and with assets of some \$4,000,000,000. But in spite of the immense power of such huge aggregations, each business concern plays a comparatively small part in the total picture of organizing and coordinating the use of all economic resources to satisfy human wants. It is still on a cooperation that is largely unorganized by any conscious effort that each individual living in a private enterprise economy such as that of the United States depends for the very means of life. Most of us take this unorganized cooperation for granted, just as we take for granted the coming of the seasons. But careful observers have often been impressed with the remarkable efficiency of unorganized cooperation.

Bastiat, a noted economist of almost a century ago, remarked,

On entering Paris which I had come to visit, I said to myself—Here are a million of human beings who would all die in a short time if the provisions of every kind ceased to flow toward this great metropolis. Imagination is baffled when it tries to appreciate the vast multiplicity of commodities which must enter tomorrow through the barriers in order to preserve the inhabitants from falling a prey to the convulsions of famine, rebellion, and pillage. And yet all sleep at this moment, and their slumbers are not disturbed for a single minute by the prospect of such a frightful catastrophe. On the other hand, eighty depart-

ments have been laboring today, without concert, without any mutual understanding, for the provisioning of Paris.¹

We have pursued at some length the central theme that the function of an economic system is the organization of a society in such a way as to allocate its scarce resources to the various uses preferred. In the performance of this function a given society develops a variety of ways of proceeding, ways which differ from one society to another. But in all of them there are certain things which must be done in some way or another. Be it America, or Russia, or Timbaktu, there must be (1) some way of determining what will be produced, (2) some way of getting resources into action to turn out these products, and (3) some way of getting the resulting products into the hands of consumers.

In the simplest of societies all of these things may be done within a single family unit. The problem of getting the wishes of the family members translated into economic action is a relatively simple and direct one. Relatively simple also is the organizing of the productive activity itself and of the division of the resulting product. The tasks are allotted by custom or by the manager of the household; the resulting products are divided among the members in the same way. Thus the family unit which directs production is the same as the family unit which consumes what its members have produced.

In contrast, where there is more than one family there is a likelihood of some interdependence between the different families. One family group develops its specialties and exchanges them for the special products of others. The consuming unit ceases to be identical with the producing unit. Even in such small communities as those of early colonial America, the benefits of specialization and exchange were evident. Although each family was to a considerable extent self-sufficing, operating its own farm, yet each tended to develop a special skill. In addition to farming, the blacksmith specialized at smithing, and became proficient at it; the miller specialized in milling, and became proficient at it. The blacksmith did a good business because his neighbors found it more advantageous to pay him for his skilled services than to shoe their own horses; the miller was able to mill flour better than each family could for itself. Each tended to produce that good or service for which he was especially well fitted and trained, and each exchanged part of his product for the products of others. Through this process all became better off than anyone could have been alone.

As soon as an important part of the organization of consumption activities is separated from an important part of the organization of production, new problems arise. These problems begin when some families specialize on some things and offer them in exchange for the special products of other families. How many blacksmiths should there

¹ Quoted from Henderson, H. D., *Supply and Demand*, Harcourt Brace and Co., New York, 1922, p. 5. The student will find the introductory chapter of this little book an excellent introduction to the study of economics.

be in a certain area? How many millers? How much flour should be milled and who should get it? The process by which these questions are answered is the process of exchange. If there are too many blacksmiths, they soon find that they do not have enough to do; people will not offer much of other things in order to get the services of a blacksmith. Some of the men who would otherwise have been blacksmiths decide to become millers instead. It is in such simple processes that we find the seeds of the unconscious organization of economic activity so necessary to the complex system of the present day. The larger the proportion of productive activity specialized and the larger the proportion of consumption obtained through exchange, the more complex our economy becomes. The more complex that economy, the greater the likelihood that it may get out of gear, the greater the danger that there may arise serious maladjustments. With such developments the need for understanding economic processes becomes increasingly urgent, and the attaining of that understanding becomes simultaneously more difficult. It is the task of the present chapter to undertake a broad preliminary survey of the ways in which economic processes are organized in a modern society.

The Advantages and Limitations of Specialization

Relation between specialization and exchange

Specialization and exchange appear hand in hand. Today we are all familiar with the assembly line techniques perfected in huge automobile plants, where each person does only one very small task but does it with the highest degree of skill and rapidity. Throughout society this specialization obtains. Skilled machinists grind metal parts to thousandths of an inch. Stenographers take shorthand and type at the rate of many words per minute. Accountants record the details of involved financial transactions. Bakers mix huge batches of dough to an exact consistency, making bread to be distributed by truck drivers, who do nothing else. And of course this specialization and division of labor among human beings is even less than the amazing specialization and division of labor now built into machinery. It is necessary to see in operation an auto assembly line, or an automatic bottling machine, or an automatic card punching and sorting machine to appreciate the wonders which mechanical genius has wrought.

But none of these efficient, specialized persons, with the possible exception of the farmer, could subsist more than a very brief period if he were suddenly to be thrown completely on his own for support. The genius who created the automatic sorting machine lives only by the very roundabout exchange of the products of his creation for the wheat of the farmer, the auto of the manufacturer, the marketing services of the storekeeper. Society obviously gains greatly in production by detailed division of labor; but such division of labor can exist only so long as exchange of the goods so produced continues freely. The auto mechanic could not afford to specialize on repairing motor vehicles unless he were

sure that others would produce the multitude of things he wants—food, shelter, clothing, amusements—and would exchange them for his repair services on their automobiles. The merchant could not afford to specialize in making available a wide range of groceries in desirable surroundings unless he were sure that others would produce the things he wants and would exchange them for his merchandising services. By the combination of our vast natural resources and our amazing system of technological efficiency and division of labor, we have lifted our plane of living to levels undreamed of a few short decades ago; but this whole upward movement depends also on the system of exchange that is essential to any specialization and division of labor. Man lives by cooperating with his fellowmen, and the more effective this cooperation the higher is his plane of living.

The advantages of specialization and exchange

Specialization of individuals in different occupations, of business enterprises in limited lines of products, of geographic areas in particular agricultural and industrial products arises in response to the opportunities such specialization offers for bettering the economic positions of the people involved. No conscious outside force has been directing the development of our economic organization. Like Topsy, it just grew. In colonial times the miller specialized because he found that the income of goods and services that he could earn in this manner was larger than that which he could obtain by attempting to be entirely self-sufficient. The farmers of Southern California specialize in growing fruits which they exchange for Iowa pork, and in this way both the Iowans and the Californians can attain higher planes of living.

Since specialization and exchange are two aspects of the same process, it is difficult to distinguish between the factors that may lead to one and those that may lead to the other, between the advantages of one and of the other. We may ask what the advantages are in developing different capacities and skills among different individuals and groups, and we may ask what the advantages are in exploiting these differences by specializing our activities and exchanging the products of these activities. In examining this question we shall consider three primary propositions: (1) Specialization of occupations of men and of nonhuman resources makes possible greater efficiency in the performing of each separate task. (2) Specialization within each business enterprise makes possible more efficient organization of production processes; the same thing is true of specialization as between different business enterprises. (3) Planes of living may be raised by personal and geographic divisions of function that focus activities on those things for which each individual or group has the greatest "comparative advantage."

1. The first of these propositions requires little elaboration. It is obvious that concentrated efforts on particular activities lead to greater skill. A man who tried to become a carpenter, doctor, farmer, stenographer, and engineer all at once would probably be inferior at all of these

occupations. This fundamental fact is expressed in the popular maxim "To be jack of all trades is to be master of none." This fact is no new discovery of our time or generation; it is recognized even among the simplest of societies.

2. The increases in efficiency made possible by specialization within and between different business enterprises is due in part to the greater efficiency of individuals specializing in the attaining of particular skills. It is also due to the saving in time made possible by limiting the number of tasks performed by any given individual or machine. The waste of time in the use of both men and machines which would result if there were frequent shifts from one task to another is obvious; it is such waste that is eliminated in the smooth organization of the modern "streamlined" factory. Moreover, modern inventions in technology as well as in the organization of production increase the advantages of specialization in many production activities. The analysis of complex tasks into separate relatively simple parts is a first step in the development of machines to displace human labor. The appearance of a machine is thus likely to lead to a greater division of tasks. And machines themselves are frequently highly specialized. Finally, the gains in organizational efficiency and in the use of machine processes within particular enterprises are matched in some respects by the efficiency gained from specialization between different firms. Modern mass production turns out machines to make machines to make machines that finally make consumers' goods. Frequently little in common exists between the producing of the machines and the producing of finished products with these machines. A division of function therefore arises between the business enterprises producing machines to produce shoes and the shoe manufacturers, or between the producers of machines to make cloth and the cloth manufacturers.

3. The basic argument for free trade between different people and different regions rests on what is frequently termed the "law of comparative advantage." The qualifications of the various individual members of a society differ; so also do the productive capacities of different geographic areas. These differences may be due to variations in native capacities and natural resources; they may be due to training and knowledge of particular skills. Wherever such differences arise there is an obvious advantage in specialization and exchange. This is clear in both occupational and regional differentiation.

If one man is a good carpenter but a very poor cook, and the other is a very good cook but a very poor carpenter, it is clear that both can gain by exchange. But the advantages of occupational specialization are not confined to cases in which one man excels in one task, another in some other task. It may be that some men are in general more proficient than others, whatever the tasks they undertake; but their advantage in some tasks will be greater than in others. It may be that Mr. Smith would be a better typist or bookkeeper or carpenter than Mr. Jones and Mr. Robinson; but his advantage over Mr. Jones is even greater if their abilities as business executives or perhaps as doctors or engineers are

compared. He will do better to devote his time to engineering and leave someone else to do the typing, even though he might also type better than the stenographer he hires. He has a greater "comparative advantage" in engineering, and both Mr. Smith and society will gain by his specialization on engineering and his delegation of typing to others. Mr. Robinson, on the other hand, is less at a disadvantage in typing than in engineering or business administration. By specializing in typing he can obtain a higher income than he could get if he were to attempt to run a business or become an engineer. Typing is more advantageous to him than engineering as an occupation, his "comparative advantage" (or comparative least disadvantage) is in typing.

The advantages of regional specialization and trade are as evident as the advantages of occupational specialization. Areas rich in mineral resources, fuel, and water power can advantageously specialize in industrial production, exchanging their products for the outputs of areas rich in good crop lands, and for the rubber and fruits of the tropics. Thus the Northeast in the United States exchanges manufactured products for the agricultural products of the Midwest and the South. The United States buys rubber from the East Indies and bananas from Central America; it sells wheat and typewriters and agricultural implements abroad.

It is obvious that areas blessed with deposits of rare minerals will trade these for other products from other regions, that tropical lands will trade tropical fruits for textiles not readily available close at hand. But the advantages of interregional trade also go further than this. A particular area may have an advantage over some other area in a long list of products, but it may specialize on some of these only and import the others. It does this because its "comparative advantage" in some of these products is greater than in others. Thus Iowa has better land for growing both corn and wheat than has Kansas, but it specializes on corn and imports wheat from Kansas because its advantage in corn growing is greater than in wheat growing. Iowa has a "comparative advantage" in the growing of corn, Kansas in the growing of wheat.

Limitations of specialization and advantages of diversification

Just as there are advantages of specialization, so there are also limitations to these advantages. Unless the quantities produced are large, a refined specialization of tasks is impracticable; the well-known activity of the man putting a single bolt on a Ford car is efficient only because so many Ford cars are produced. Extensive specialization requires large markets for the goods produced. Moreover, the greater the degree of specialization the more complex the problem of coordinating the activities of the different individuals within an enterprise and the more vast the problem of organizing trade and exchange between the many different sectors of the economy. Extensive specialization in manufacture and farming requires also an elaborate development of marketing facilities,

such as transportation, finance, wholesaling. This is costly. And along with all these things comes insecurity far greater than in a simpler and more self-sufficing organization of economic activity.

Diversification may be advantageous if tasks are so related that the performing of one contributes to efficiency in the performing of others. To "know more and more about less and less" is not always the most efficient procedure. Diversification in the pursuit of related occupational skills is clearly important for those in administrative and directing positions in economic life; it may also lead to greater efficiency in many crafts and white-collar jobs. The ability to "fit in" with some degree of elasticity is a good recommendation. Within a business enterprise diversification may be a part of the economical use of by-products of technical processes; it may be a direct outcome of coordinated industrial research; it may be a response to consumer demands. It is difficult to imagine anything much more diversified than the offerings of a large modern department store. Diversification in a business enterprise gives the enterprise some degree of stability in the face of changing market situations. And diversification within a region or a nation provides some protection against the insecurity of erratic and drastic changes in outputs and markets for single products.

Specialization and Exchange in the Modern American Economy

Factors encouraging specialization in the American economy

The development of a complex highly specialized mass production economy in the United States is not a matter of historical accident. This country has had the advantages of access to a wealth of natural resources and to a wide and free area of trade. When we speak of a "big market" for a product we mean that there are many buyers of that product. Where there are large populations in a position to buy there are large market opportunities. The United States grew rapidly in both population and area. The populations of geographically distant places were brought together by the development of a system of transportation previously unknown. Whereas in Europe the market was split by restrictions on trade between the different nations, in America we were more fortunate. The founders of the nation provided that there should be no trade barriers between the states, and despite the fact that some restrictions to interstate trade have appeared, we have developed essentially as one vast and free economy. It is this fortunate conjuncture of circumstances that provided the basis for making the United States the leading mass-production nation of the early part of the twentieth century.

Occupational specialization in America today

The complexity of occupational specialization in America today defies full comprehension. The major sectors of economic activity of gainfully

occupied persons over 16 years of age are shown for the years since 1870 in Figure 2—1 and Table 2—1. The data for the last decade, while not yet available in final form, indicate generally a continuance of the trends shown in this chart, except for a leveling off of the relative position of

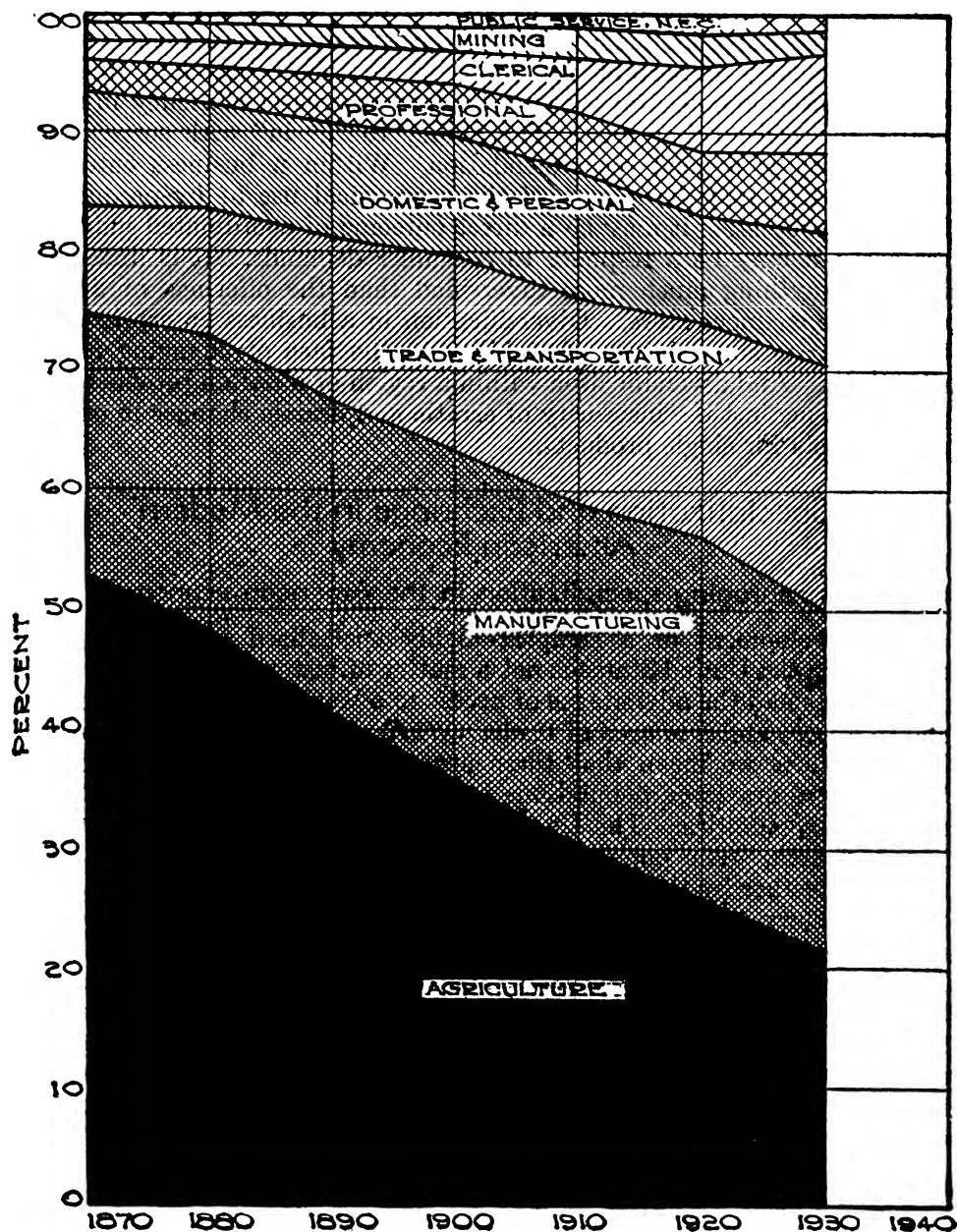


Fig. 2—1 Proportion of gainfully employed over 16 years of age by major occupational groups, 1870—1930.*

* Reproduced from *The Structure of the American Economy*, National Resources Committee (Washington, 1939), p. 65.

manufacturing, and a somewhat more rapid growth in trade and transportation and in public service. A more detailed breakdown for the year 1935 is given in Table 2—2, but even these are broad classifications, within each of which is included a heterogeneous combination of many particular occupations and skills. A detailed classification would run into tens of thousands of different occupations, ranging all the way from

TABLE 2—1 *

NUMBER OF PERSONS 16 YEARS OF AGE AND OVER IN SELECTED OCCUPATIONAL GROUPS, 1870-1930 ^a

Occupation Group	(In thousands)						
	1870 ^b	1880	1890 ^c	1900	1910	1920	1930
Agriculture and allied occupations	6,428	7,830	8,973	9,802	^d 10,872	^e 10,524	10,242
Mining	172	252	388	576	947	1,083	983
Manufacturing and mechanical industries	2,674	4,033	5,743	7,537	10,253	12,425	13,790
Trade transportation	1,104	1,741	2,969	4,445	6,223	7,360	9,963
Clerical occupations	206	330	543	781	1,635	2,952	3,935
Domestic and personal service	1,168	1,437	2,133	2,726	3,805	3,605	5,448
Professional service	338	543	880	1,196	1,727	2,203	3,110
Public service not elsewhere classified	73	107	185	260	382	642	692

* Source: *Recent Social Trends in the United States*, Whittlesey House, 1936, Table 6, p. 21. Figures differ slightly from those in Census reports; adjustments have been made to obtain comparable figures.

^a The figures of this table are the result of an attempt to derive presumably comparable series of figures from the successive occupation censuses. The figures will not in all cases be found in the census reports. In a few instances, estimates have been made by dividing figures for combined occupational groups contained in reports of the earlier censuses, and in other instances by combining separate census figures for later years. The composition of some of the major groups differs somewhat from that of similarly designated categories in the recent census reports, because the recent census categories could not be carried back to the earlier years. Other estimates and adjustments in the published census figures have been made, the more important of which are indicated in the following footnotes.

^b Figures for total occupied and for large occupation groups in 1870 are adjusted for the probable deficiency in the population enumeration in that year.

^c Figures for total occupied and for the larger occupation groups in 1890 are estimated from published figures for gainfully occupied persons 15 years of age and over.

^d Figures for farm laborers in 1910 are adjusted for supposed overenumeration of women and children in agriculture. The adjustment probably leaves the figures for agricultural laborers in this year still too high.

^e Figure for paid farm laborers in 1920 adjusted for probable underenumeration due to the date of the census of that year.

common unskilled labor at ditch-digging to the most highly trained research technicians, doctors, and "big businessmen."

This occupational specialization has a profound effect on the lives of the people of America. It is a far cry from the more intimate community of colonial days. The variety of interests which was earlier associated with diversity in the job is for many individuals now obtainable only through the use of leisure time. At the same time, business executives may be faced by a variety of situations and problems which far outruns anything known in earlier generations. And finally, the insecurity of the individual, and the increased responsibility of the public for providing security have in recent years led to profound changes in our economic ideologies and our concepts of the role of the state in the life of the people.

TABLE 2-2*

THE AMERICAN ECONOMY 1935
EQUIVALENT FULL TIME EMPLOYMENT—41,411,000
(In thousands)

Total Agriculture	9925	Fertilizer	17
Total Minerals	664	Coke	17
Petroleum	109	Metal Mfg.	2229
Nonmetallic	44	Iron and Steel ..	375
Bituminous Coal	353	Nonferrous Metals except Cop-	
Anthracite Coal	78	per	32
Metallic	80	Copper	10
Total Manufacturing	7204	Other Iron and Steel	351
Food, Tobacco and Leather		Nonferrous Metal Prod.	238
Products	1195	Automobiles	388
Dairy Products ..	49	Other Trans. Equip.	93
Meat Packing	117	Electrical Machines ..	180
Flour Milling	26	Other Machines ..	577
Canning	130	Printing and Publishing	305
Tobacco Mfg.	90	Book, Music and Job ..	127
Sugar Refining	26	Newspapers and Periodicals ..	119
Leather	51	Allied Industries	59
Baking ..	218	Non-Wage Earners Mining and	
Other Food	171	Manufacturing	1386
Confectionery	60	Bulk Distribution ..	
Shoes	202	Wholesale Trade ..	1350
Other Leather Prod.	58	Business Service ..	252
Textiles	1739	Retail Trade ..	4970
Wool and Worsted	158	Construction ..	719
Cotton Textiles	369	Services to the Consumer	6563
Silk and Rayon	126	Public Education ..	1152
Rayon Yarns	51	Private Education	207
Finishing	75	Professional Service	1337
Knit Goods	220	Personal Service ..	1488
Wearing App. Fabric	574	Domestic Service ..	1823
Other Textile Prod.	166	Recreation and Amuse.	232
Pulp and Paper	127	Miscellaneous Serv.	324
Lumber and Millwork	304	Services to the Economy	6347
Rubber Tires	57	Federal Gov't Except Post Office	790
Other Rubber Prod.	58	State and Local Gov't except	
Paper Products	109	Public Education	1241
Furniture and Related Industries	134	Banking and Finance	1097
Misc. Lumber Prod.	137	Post Office	260
Miscellaneous Mfg.	249	Telephone	292
Chemicals	188	Telegraph	70
Paints and Varnishes ..	28	Electric Power	231
Clay Products Ex. Pottery ..	45	Gas	51
Pottery	29	Rail Transport	1113
Marble, Granite and Other		Auto Transport ..	755
Products ..	72	Other Transport	447
Petroleum Refining	77	Miscellaneous	2029
Glass	67		
Cement	21		

* Source: National Resources Committee, Industrial Committee, *Patterns of Resource Use, 1935*, Table II.

Regional specialization in America today

Daily newspapers are filled with comments and news involving differences in economic interests of different parts of the United States—the textiles of New England, the coal and steel of Pennsylvania, Ohio, and Illinois, the cotton and tobacco of the South, the corn and wheat of the Middle West, the lumber of Washington and Oregon, the range cattle of the Rockies, and the fruit of California and Florida.

One way of measuring the geographic distribution of production is to record the numbers of people employed by locations and activities. The authors of *The Structure of the American Economy* classify the broad categories of productive activity into three groupings according to the nature of their geographic distribution as measured by numbers employed. These groupings are indicated in Table 2—3.

TABLE 2—3 *

LOCATION OF PERSONS EMPLOYED IN RELATION TO RESOURCES AND CONSUMERS, 1935

	<i>Number of persons (thousands)</i>
Located close to resources:	
Agriculture	9,925
Fishing	65
Mining	745
Manufacturing †	743
Located close to consumers:	
Services to consumer	6,563
Services to the economy	6,346
Services to business	252
Retail trade	4,970
Construction	719
Manufacturing	887
Relatively footloose:	
Wholesale trade	1,350
Manufacturing	6,881
Miscellaneous	1,965
TOTAL	41,411

† Including sawmill and timber workers.

* Source: *Structure of the American Economy*, p. 36.

Specialization in the production of a single consumers' good

To complete the production of any product and get it to the consumer in its finished form involves the use of a wide range of human and material resources, reaching through all the complex interrelationships of the economic system. Strictly speaking, no commodity is produced without the cooperation of productive agents going back to the very beginning of time, since virtually all of our present productive capacity depends to some extent on productive activity of preceding periods.

Such an apparently simple product as the can of corn which you buy at the corner grocery illustrates this. The corn itself represents in the

first instance the cooperation and specialized activities of the farmer, his farm machinery, his seed, and his available natural resources. But the machinery represents the activities of farm implement manufacturers, of steel mills, and of transportation facilities. Back of these lie the mining of coal and iron ore, and so on. The farmer's home and barns represent the cooperation of carpenters, plumbers, mill hands, lumber-jacks. Turning to the tin can, we can trace the thread directly back through the transportation facilities to the can manufacturer, and beyond him to the steel mills, the manufacturers of his machinery, equipment, plant, and so on. The contribution of the corner grocer rests on his own equipment, his workers, his buying connections. Superimposed over all this physical activity is the financial activity, which plays an important role in facilitating the physical production and the distribution of goods and services to consumers. Banks, insurance companies, stock exchange brokers, wholesalers—all these and many other specialists have contributed to the production of your can of corn. To trace fully the range of cooperation going into the provision of even this small article would take a full volume if confined to activities of a current time period; it would be an unending task if all the connections with past productive activity were to be followed out.

The Functions of "Middlemen" in a Complex Exchange Economy

An American college student awakens in a bed probably of steel made from iron mined in the Mesabi range of Minnesota and processed in Gary, Indiana. He throws back covers made of wool from sheep raised by a shepherd in New Zealand or Australia, spun and woven in Leeds, England, or perhaps in Manchester, New Hampshire. His sheets are made of cotton cultivated in the Southeast, sold in New York markets, and spun and woven in the mills of Fall River, Massachusetts. He washes with soap made from vegetable oils from the tropics. He puts on shoes which are made from the hides of animals grazed on the open ranges of the western mountain states, processed in Chicago, cut, sewn, and finished in St. Louis, Missouri. The heels are of rubber grown in the Malay Peninsula and shipped to America in boats built in Holland or England out of iron from the Ruhr valley and wood from Scandinavia. He dons pants and sweater, again probably of wool from the other side of the world, processed into cloth as was his blanket, cut, styled, and sewn in factories in New York or Chicago. He straightens a tie which may be of pure silk from the Orient, or of rayon made out of cellulose from plants grown in the Southern states and processed by Du Pont. Glancing hastily at a watch made in Elgin, Illinois, out of silver from Nevada or gold from Colorado, he hurries to breakfast. He drinks coffee imported from Brazil, eats oranges from Florida or California, cereals made from grains grown in the prairie states, and sugar grown in Cuba and refined in Baltimore. He lights a cigaret made of tobacco grown in North Carolina or Kentucky and packed perhaps in Virginia, with a

match produced in Ohio. Finally he hurries to class where his instructor hands him an assignment mimeographed on paper from Erie, Pennsylvania.

None of these items reflects a particularly unusual plane of living. They are parts of our ordinary everyday lives. And how do we get all these things together for the use of this college student when and where he wants them? Who is responsible for shipping the coffee from Brazil, storing it, insuring it, finding who has it and who wants it and in what amounts? Who finances the producers while they were waiting for our student, the ultimate consumer, to buy these goods? Who makes it possible for the buyers of cotton for further processing to know what they are getting, and how is this done? Who takes the chance that there may be damage to the stored wheat or that consumers may decide they would rather have rayon than silk ties? Who gets the things this student wants to the place where he finds them?

The people who perform all these services for us are often called "middlemen." The functions that they perform in getting to people the goods they want, when and where they want them, are called "marketing functions." The most frequently used classifications of marketing functions set them up somewhat as follows:

- A. Exchange
 - 1. Selling
 - 2. Buying
- B. Information
 - 1. Market news
 - 2. Description and grading
- C. Physical supply
 - 1. Transportation
 - 2. Storage
- D. Auxiliary services
 - 1. Financing
 - 2. Risk-taking

Exchange

The function of selling involves the work of finding buyers for the products of farmers and manufacturers, discovering where the demand for a good lies and directing it into these channels. It includes the providing of information to potential buyers so that they may know what to choose. This function is a part of the whole process of production. Insofar as it is carried out honestly and insofar as information is given fully, it is a service which will contribute to the allocation of resources in accordance with consumers' preferences. Insofar as the supposed information is misinformation, or insofar as the finding of buyers is transformed into the coercion of buyers, it will work in the opposite direction. It is evident, however, that in a complex exchange economy the function of the seller is a necessary one.

The function of assembly (or buying) is the service of acquiring legal title to a quantity of goods in order to have them accessible to either the ultimate consumers or to other middlemen or processors. It is not the physical assembling of goods. This would be transportation. It is merely the service of providing a clearinghouse to which buyers may turn. This is what is done on the big commodity exchanges when speculators buy wheat or cotton which is never actually delivered to them but is again sold to someone further along in the line from producers to consumers. Such a function is also performed by a wholesale house or a department store when either seeks out the sources from which the various items demanded by its customers may be obtained. The middlemen engaged in assembly are carrying the expressed or anticipated preferences of consumers back through the economy to find for them the goods they wish.

Information

Market news concerning such things as prices and the flow of goods is the guide to important decisions on the part of producers engaged in all the many activities by which goods and services are finally made available to consumers. It is on the basis of such information that manufacturers plan their future production programs, that buyers and sellers decide where, when, and in what quantities to buy and sell. Such information forms the basis of decisions as to where goods should be shipped and when they should be held in storage. The greater separation between consumers and producers has made the providing of market news much more important than in a simpler society. Not only is there greater specialization, but there is the necessity of planning further into the future. Increased information reduces risk and enables producers to anticipate consumer demand more accurately. It reduces waste by avoiding the production of unwanted goods and by eliminating charges for financing or for unnecessary storage. Those engaged in providing this service are key men in passing on through the economy knowledge or estimates as to what consumers' preferences will be and how they may be most economically satisfied.

The grading and description of products, like the providing of market news, has taken on much greater importance with the widening gap between consumers and producers. With each exchange, and with every order for the future, arises the problem of specifying the quality and the characteristics of goods to be exchanged. Grading involves the setting up of classifications or standards by which goods may be described. It is a necessary implement for providing market news. Much waste can be avoided if buying and selling can take place on the basis of more extensive knowledge. The importance of this fact is appreciated by most producers, and goods often are graded in raw-material markets and wholesale markets. It is unusual, however, to find goods graded in the retail markets. The explanation lies in certain major differences between the retail and other markets. In the first place goods are less differentiated

in the producers' market (where both buyers and sellers are producers) than they are in retail markets. They are therefore more easily described in standardized terms. In the second place, there is a difference in the degree of specialization of the buyers. Commercial buyers specialize on a few things and come to know how to judge them; consumers buy a wide range of goods and become informed concerning only a few of them. There can be no doubt that knowledge is very imperfect in the retail market where consumers' preferences are first expressed, and errors made there are carried on back and reflected in the plans of producers.

Physical supply

The functions of physical supply, which include transportation and storage, are obviously essential and the nature of their service in the economy is easily understood. In a society which draws on products from countries all over the world transportation is a vital connecting link. In even the simplest economy storage is necessary. The fact that agriculture is seasonal makes this inevitable. The increased complexity of our exchange system has increased the need for storage in other situations. It is required if there is to be continuous operation in manufacture when demand is seasonal. It is also necessary to carry a certain amount of inventories in order to be prepared to meet the demands of consumers. Goods are sometimes stored locally in order to transport them in carload lots rather than in small shipments. Clearly these services are a part of the process of getting the goods consumers want to those who want them, when and where they are wanted.

Auxiliary functions

The auxiliary functions exercised by middlemen facilitate all phases of production, including agriculture, manufacturing, and marketing. Among these functions is financing. Considerable time elapses between the initiating of a productive process and the sale of the good or service which is the end product of this process. Meanwhile the resources used in the process must be paid for as used. "Working capital" is necessary in the interval between these payments and the sale of the goods. The producer must either do his own financing, using his own funds and foregoing the interest he might receive elsewhere, or he must have recourse to a bank or other financial agency which will make advances on the security of the good in process. In either case a cost is incurred for this financing.

A function frequently closely associated with financing is that of risk-taking. Some risks can be insured against, and the specialist in the providing of insurance will find his business as stable as do others. Thus life insurance, insurance against fire and theft and hail are all services provided by specialists. Risks can be insured when they can be "averaged out" so that though it is impossible to predict what will happen to any particular person or house, one may have a reasonable degree of assurance that a certain number of houses will burn, jewels will be stolen,

people will die, and the costs ensuing may be paid out of the insurance premiums of the larger group.

But in a society full of uncertainties, there are many risks not so easily provided against, risks which are more capricious and involve the hazards of changes in consumers' demands, in the costs of raw materials, and other unpredictable factors. A businessman may carry all these risks himself or he may seek to shift them to the shoulders of a "speculator," who is a specialist in such risk-taking. Despite the many abuses which have given "speculation" such a bad name, it is an important and constructive function which must be performed by some one in modern society.

The quantitative importance of middlemen

It is not always easy to separate from other activities such as manufacture and agriculture the activities which would be regarded as the functions of middlemen. A rough estimate is available, however, in the data on occupations presented in Table 2—2 on page 28. Of a total of approximately 41,411,000 people employed in 1935, about 10,039,000 were employed in sectors of economic life serving primarily the functions of middlemen. This includes 6,602,000 in trade (buying and selling at wholesale and retail), 2,315,000 in transportation, and 1,097,000 in finance. It omits the marketing and communications functions of governmental agencies as well as those marketing services which manufacturers and others perform for themselves. The great increase in the importance of middlemen is also evident in the jump in the relative importance of "trade and transportation" since 1870, as shown in Figure 2—1 and Table 2—1.

Various other estimates of the quantitative importance of the functions of middlemen are available. One of the best² indicates that for every dollar of manufacturing costs, there is 85 cents in the costs of retailing, wholesaling, and marketing services provided by manufacturers. A roughly similar picture is presented by a breakdown of the dollar which the consumer spends for food products. Of this dollar, approximately 41 cents goes to the farmer, 20 cents to processors, 5 cents to transportation, 9 cents to wholesalers, and 25 cents to retailers—that is, roughly 60 cents to farmers and processors and 40 cents to marketers.³

Are there too many middlemen?

There have been frequent complaints in recent years that "distribution costs too much," that somebody steps in between the farmer and the consumers and walks off with the "velvet." While there may be waste in providing the services of middlemen, just as there may be in the operation of farms and factories, it does not therefore follow that middlemen

² Galbraith, J. K., and Black, J. D., "The Quantitative Position of Marketing," *Quarterly Journal of Economics*, Vol. 49, p. 402.

³ National Resources Committee, *The Structure of the American Economy*, 1939, p. 69.

should be especially subjected to attack. The naive individual who thinks that the only real production is manufacturing or mining or agriculture is under a serious misapprehension. We can have efficient, highly specialized agriculture and manufacturing only if we have also an organization of services for getting products from those who first produce them to the ultimate consumers. Wherever there are advantages of specialization there is also a call for the services of middlemen. The more elaborate and complex the specialization in physical production, the more costly the job of transporting and exchanging and financing the goods produced. Thus the costs of marketing are a factor restraining the development of ever more and more division of labor between man and man, firm and firm, area and area. To think that merely by establishing a cooperative the consumers or the producers, as the case may be, will then be able to get the supposed "velvet" previously "looted" by middlemen is to court disaster. The functions which these middlemen perform must be performed by someone, and they cost something. They are quite as much a part of the production of economic goods and services as are agriculture and manufacturing.

We might as well ask, "Are there too many farmers?" or "Are there too many manufacturers?" as to ask, "Are there too many middlemen?" The answer can be given only after a thorough examination of the way the system is functioning, of where particular kinds of waste exist, of the extent to which consumers' preferences are finding expression in productive action, of the distribution of incomes.

A Simplified Model of a Free Price System

Awareness of the astounding complexity of the modern exchange economy only serves to re-emphasize the question asked earlier: What sets up and directs this remarkably complex organization? Somehow it must be determined what goods and services are to be considered most important; then they must be produced; and last they must be distributed to the people who want them—the consumers.

To grasp the magnitude of the problem involved, suppose for a moment that you are made complete economic dictator, and that for some unfortunate reason you have not the slightest idea how much of everything is now being produced. What you would have to do is tell each worker in the economy just what job to do, determine just how every other productive resource in the economy shall be used, and decide just how much of each commodity and service shall be produced, and at what prices these things should be sold—and all this so that the amount produced of each good and service will be equal to the amount demanded and so that no productive resources will be involuntarily unemployed! How, for example, would you decide whether Mr. X ought to spend his time grinding fine edges on a machine tool later to be used in the manufacture of tractors later to be used in the production of corn, later to be fed to hogs later to be eaten as pork, or whether he ought to spend his time

making fine cabinets, or whether he ought to spend his time running a locomotive, or whether he ought to spend his time operating a corner grocery store? What would you do about the apparently minor problem of how many sheep to raise, when each sheep yields both wool and mutton, the demands for which are almost completely independent? Wool is used for clothing, blankets, rugs; mutton is used for dinner. How are you going to make them come out even, even if you have somehow decided on the total number of sheep to be raised? How would you decide who is to have the tractors and corn and cabinets and mutton and wool once they are produced? These are the merest suggestion of the problems to be faced—and it matters not what sort of economic system exists, be it capitalist, socialist, fascist, or what not. Somehow these decisions have to be made.

In the American economy we have depended primarily on a relatively free price system to provide (1) indexes of consumer preferences as a basis of determining what to produce, (2) the incentives to get these goods produced, and (3) a mechanism for dividing the resulting product among the individual members of the society.

How the free price system registers consumers' preferences

The pricing process provides a way of registering consumer preferences through consumer money expenditures on various goods and services. The more the consumers want something, the higher the price that they will be willing to pay for it. The relative strength of consumer demand for various products is measured in the prices of the products, and it is on the basis of this pattern of prices that the price system decides, so to speak, which are the most important goods and services to be produced. It should be noted at the outset, however, that the consumers' demands which register themselves through the price system are demands expressed in money offers. No matter how badly a poor family may want a Buick car and a fine home, its desires become effective in the market only to the extent that they are backed up by the ability to pay for such things. The price mechanism registers consumers' demands as the controlling factor in deciding what shall be produced, but the only consumers' preferences which are counted are those expressed in money offers. Therefore the rich man has far more power to determine what things are "important" to produce than has the poor man. Prices reflect not how much consumers "need" goods and services, but how much they are willing and able to pay for them.

How the free price system gets desired goods produced

The desires of individuals to maximize their incomes lead roughly to the production of those things registered through the pricing process as most important. These incomes appear in the form of profits, wages, rents, and so on. Business enterprise is operated for a profit, and businessmen will seek to produce the good or service which promises to yield the highest profit. The greater the consumer demand for anything, the more the sellers can get for it, and the more eager will businessmen be to

produce it. Thus the pricing process acts, however imperfectly, as a device for registering consumers' preferences; and the desire to get as great an income as possible simultaneously calls forth production of those things most in demand. Likewise, the profit motive encourages the most efficient (low cost) production, since the lower the cost at which a businessman can put his product on the market the greater will be his profit.

In his efforts to produce the things which promise the greatest profit, the businessman employs productive resources—men, machines, land, and so on. He pays to laborers or owners of other productive resources a "price" for the services they contribute. Owners of productive resources will seek the employer who offers them the highest reward ("price"). Thus the man running a business is essentially a link between consumers and productive resources; his social function is to organize productive activity and channel the use of productive resources toward fulfillment of consumer demands. Profits provide the incentive which induces him to carry out this function.

Any interference with free access to profit opportunities is an interference with the operation of this free price system. Thus if government taxes some industries and pays subsidies to others it will change the allocation of resources; the taxed industry will become relatively less attractive, the subsidized industry more so. And if some private groups get exclusive control over the production of certain commodities so that they are able to keep other businessmen out, this also is an interference with the operation of the free pricing system and will alter the allocation of resources by preventing responses to profit opportunities from taking effect.

How the free price system distributes the products of economic activity

There are two steps in the process by which the free price system distributes finished goods and services among the many consumers who want them. First, it determines what different individuals will receive as money incomes; these incomes are the "prices" paid for the productive contributions made by different individuals in their capacities as laborers, owners of machines, of land, of buildings, and so on. How much income each of us receives largely determines how much of the goods produced we can buy. Second, the price of each commodity is bid up until the buyers least able or willing to pay for it are eliminated. So the goods produced are distributed. Of course this does not necessarily mean that the buyers least able to pay are eliminated completely. More commonly it means that they can afford only a few units of a good at the price established, while higher income groups can afford more; though in buying some goods, such as fur coats, and crystal vases, poor persons are eliminated completely.

The circular flow of economic activity

The web of interrelated actions involved in performing the three basic functions of the free price system now begins to appear. The price

system gives us our incomes as the prices of the productive services we offer; it registers our preferences when we, as consumers, spend these incomes; and it calls forth the production of those finished products for which the strongest consumer demands have been registered, using in the productive process the productive services, payment for which in turn gives rise to further incomes and consumer demands.

It is important that this continuing circular flow of economic activity be thoroughly understood. Figure 2—2 pictures the flow of money in one direction through the system and the counter-directional flow of productive services and finished products. Businessmen act as the intermediaries between people as sellers of productive services and as consumers of finished products. The consumers spend their incomes on the

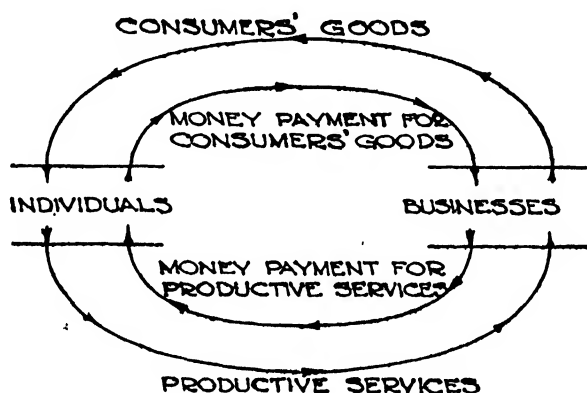


Fig. 2—2. The circular flow of economic activity.

goods sold by the business enterprises, and the money they pay moves through the business enterprises and becomes incomes once more. These money payments of businesses to individuals are rewards for productive services rendered, and these productive services result in the products that are given to the consumers in exchange for the dollars they spend. Thus there is a flow of money through the system from consumers to the sellers of consumer goods and services; and from these businessmen back to the laborers and owners of productive resources, who then spend this money once more in consumer purchases from business enterprises. Each time money is passed from hand to hand something is given in exchange for this money, either productive services or the consumer goods that they turn out. Thus we may say that people's contributions (of labor or of other productive resources) in production give rise to the incomes which permit them to live; or we may say that the funds people spend for finished products give rise to the jobs by which they earn their incomes. Both are correct and neither is more fundamental than the other. Economic activity is a continuous flow of productive services and finished products, implemented by the flow of money in the opposite direction, and guided by the price system into thousands of different channels.

The Operation of the American Economy

The simplified picture of a free price system which we have just sketched is only a crude outline of the ways in which, through the self-interested actions of thousands of separate individuals, the resources of an entire society may be productively channeled into fulfilling consumer preferences. There has probably never been a society so simple as this, a society in which all actions affecting resource allocation are based on individual initiative without government interference, a society in which there is no significant curtailing of opportunity through the development of powerful private group interests. Yet this simplification gives us a glimpse of how economic processes work in a society such as ours. We are now ready to turn to the American economy and see briefly what major modifications appear and what the major gaps are in the picture just presented.

Displacement of individual by group action

The simplified model of a free price system is most nearly approximated in an individualist economy. Such an economy is characterized by a large proportion of individual responsibility and initiative in production, and by the influence of individual consumer decisions on the allocation of resources. There may be various degrees of private control over property and job opportunities. Individuals have the legal power to decide what they want to do, where, and when, though in periods of depression there may appear to be little real freedom of choice.

The economy of America in the middle of the nineteenth century was a highly individualist economy. The decisions in the economic system were made by individual consumers and individual producers in their own particular interests. In the United States today the traditional individual initiative is being replaced in more and more spheres by group controls—by farm organizations, by corporations and trade associations, by trade unions. Individualism appears to be undergoing a transition into what we might term “groupism.” The price system is still the principal organizing mechanism in the economy, but it is in many respects not a “free” price system. This rise of economic groups is not new in the world; it is only a somewhat different form of an old phenomenon which predated the era of individualism and was for a time submerged.

There are many private interferences with the free flow of economic activity. These interferences become important when controls shift from separate individuals to organized interest groups. Such groups seek to obtain especially favorable positions and to maintain themselves in these positions by a wide variety of restrictive practices. For example, instead of a free adjustment of prices in some industries, businessmen band together to hold prices at some agreed level. They engage in elaborate and skillful maneuvers to prevent new enterprises from entering a profitable trade in competition with them. Laborers may build up strong trade unions which gain control over jobs in a particular skill or industry, and

prevent the free entry of other workers. A major task of much of economic study is to examine the conditions under which private group action supports or interferes with smooth adjustments in the price system and hence with the resource allocation and the income distribution which it would bring about.

Government in the American economy

The price system operates within a legal and social framework which supplies the rules of the game and acts as umpire in cases of dispute. The legal foundations of individualism in America are expressed in the constitutional guarantees that no one shall be deprived of "life, liberty, or property without due process of law." "Property" has been interpreted to include not only control over the use of physical things, but the expectations which a man has of future income from a continuation of present practices. Associated with these ideas was the attitude that government should be confined to a minimum of functions—the less it interfered with private actions the better.

Individualism has never gone so far, however, as to exclude government intervention altogether, nor is such a situation conceivable. Freedom for one man necessarily involves restrictions on the freedom of others. Questions arise as to how far government may or should go in restricting some freedoms in order to preserve others. Attempts to solve this problem were based in previous generations on the assumption that typically individuals made their own decisions, and served as effective checks on each other. With the increasing complexity of our economic structure and the development of more and greater power groups in labor, among businessmen, among farmers, new problems arise. In formulating and evaluating social policy private group action must be recognized as a significant element in the total situation.

Government action may be designed to influence economic adjustments at any point in the system. It may intervene in the registering of consumer preferences by providing information to consumers on the one hand, or by controlling production and imposing some kind of rationing system in the place of a price mechanism on the other. It may intervene in the processes of production in any of hundreds of ways; by taxes which penalize certain industries (as alcoholic beverages), by subsidies which encourage others (as housing), by programs to break up power group controls over production and prices (as in the antitrust acts). It may intervene both directly and indirectly to modify the income distribution resulting from the operation of the unregulated price mechanism. Moreover, some goods and services are not readily provided through the free price system. Examples are the army, navy, police, and roads. And when we desire collectively through our government to encourage very much the consumption of certain things, it may be more efficient to produce them through government action directly; this is especially true when we wish to make consumption compulsory, as in the case of public education and some public health services.

The scope of government action in economic life in America has been steadily increasing, though except for the present war emergency it apparently is still a long way from becoming a centralized economic system managed by government. Despite the increased importance of power groups displacing individual initiative, and despite the increasing scope of government action, the American economy is still rooted in individualism. It remains a private enterprise system in which economic life is organized primarily through the pricing mechanism of the market. This mechanism works very imperfectly, and requires considerable repair and overhauling. The examination of how it functions in the present setting, and the use of this information in the evaluation and the formulation of social policy are primary purposes of this book.

The Method of Economic Theory

Economic analysis is the analysis of certain aspects of social structure—primarily the use and allocation to different ends of scarce resources which are transferable. The economic system is thus itself an abstraction, a convenient setting apart of an aspect of social action in order to facilitate understanding. This does not mean that it is unreal. The very existence of economics as distinct from other branches of social science is a recognition of the fact that the total is very complex—that by separating it into abstractly constructed segments, such as economics, sociology, or political science, we hope to arrive at increased understanding of the whole. Even within the sphere that we have defined as “economic” the problems are so vast and complex that we should be lost in the woods without some further simplifying technique.

This technique is provided in the construction of “ideal types” that form the foundation of economic theory, as they do of most other fields of scientific endeavor. There is nothing especially desirable or undesirable about an “ideal type.” We use that phrase because it is a simplified model constructed in our minds, made up as an “idea.” It is not at all a photograph, or an average, of reality. It is neither something we seek to attain nor to avoid.

Let us take first a very simple illustration outside economics. There is no such thing as a two-dimensional surface, yet geometric concepts of straight lines, squares, circles, 90-degree angles, or parallelograms are of very real practical value. No one could find an actual square, yet no one would challenge the validity of the *idea* of a square. The “circle” and the “square” are “ideal types”: they are abstract simplifications of certain aspects of concrete reality. The test of the significance of an “ideal type” is this: can it be used effectively in the understanding of the concrete phenomena to which it serves as a simplifying approximation?

As used in economics “ideal types” are abstracted aspects of life, just as economics itself is an abstracted aspect of the total of social life. The model of a free price system is such an “ideal type.” There are many

"ideal types" used in economic analysis, some more frequently than others. Sometimes we talk about what would happen if there were many producers of an identical product. Sometimes we talk about what would happen if there were only one producer of a good. There are many other variations. Each is a much simplified model, with many of the variations that would appear in any concrete situation omitted. The model may approximate existing conditions to a greater or a less degree. It may be elaborated to approach more and more closely certain concrete situations, but this process leads it to become more and more complex, and hence less and less understandable. Carried to extremes each case would become a model in itself and we should be lost in the woods again.

The most feasible approach to an understanding of the economic system is therefore to begin with an intensive study of a few relatively simple "ideal types" that include the predominant characteristics of the concrete economy in which we are interested. In analyzing a particular problem we can take as a first approximation that particular "ideal type" that seems most nearly to fit the situation. The next step is to make such adjustments in these models as may render them more applicable to the particular situation under consideration.

Those models of economic theory that will prove most useful will depend on the structure of the particular society in which we are seeking to apply our analysis, and on the social goals of the society. The "ideal types" that are most useful in the analysis of an economy largely controlled by a totalitarian central government differ in many important respects from those most useful in understanding the American economy and in formulating social policy in this country. But whatever the social scene, the essential method of economic analysis remains the same. It is the method of building up simplified abstract models of aspects of economic life, models that we call "ideal types," and of improving upon and modifying these models continuously in order that they may be as useful as possible in facilitating an understanding of the complex economic processes going on around us.

CHAPTER 3

Consumer Demand

INDIVIDUALS express their preferences by spending their incomes on the various goods and services available in the markets. The actions of an individual toward any given commodity, say shoes, pork chops, or eggs, will depend on (1) his income, (2) the rating of various goods in his scale of preferences, and (3) the availability of these various goods. At any given income, with any given scale of preferences, and any given set of available alternatives in the spending of his income, this individual will buy a certain number of eggs per week at a high price for eggs, and some other quantity at a lower price for eggs. This set of possible actions toward eggs is his demand for eggs. Others will have different demands for eggs. In our profit-motive economy it is to the demands of the aggregate of consumers for shoes, pork, eggs, and so on, that producers ultimately respond in deciding to produce one thing or another. The demands of ultimate consumers for finished goods and services are reflected in demands of producers for productive resources—for example, for coal and steel, for machines, labor, and raw materials. Thus consumer demand is the logical starting point for an examination of the way society's scarce productive resources are allocated to the satisfying of the multitude of human wants included in individual scales of preferences.

The Meaning of the Term "Demand"

In much discussion, loose and inconsistent usage of the terms "supply" and "demand" is likely to invalidate the conclusions reached. As with other scientific terminology, analytical terms in economics must be precisely defined and consistently used if they are to be of assistance in analyzing problems. Otherwise nothing but confusion can ensue. In taking care to define terms sharply and to use them consistently, the economist sometimes gives a very precise definition to a term which in lay discussions is used loosely and with different connotations. Therefore it is imperative that great care be taken to see exactly what each analytical term is defined to mean and to remember to use it strictly in that sense. This warning applies to all of the terminology of economics. It is particularly important in defining "demand."

Individual demand

The demand of an individual for any given good is the several amounts of that good which he would buy at each of an associated series of prices, other conditions remaining unchanged.✓

This demand must be expressed in terms of some unit of time—the number of eggs purchased at 20 cents and 30 cents and 40 cents per dozen *per week*, or *per month*. Without such a statement of the time period involved, a statement of how many eggs will be purchased at each price would be meaningless. ✓ Thus individual demand is a list or *schedule* of the amounts which would be purchased by the individual at various prices of the good in question for some given time interval.✓

It is essential to understand that it is meaningless to speak of demand for a good as an absolute quantity, as when one says, for example, that “the demand for wheat is ten million bushels.” Obviously, under any given set of conditions people will buy more wheat at 30 cents a bushel than at \$10 a bushel. How much of anything people will buy has meaning only with reference to some particular price or series of prices. If wheat were free and obtainable without effort consumers could use infinitely more than they do, not only for food but for fuel, and for a wide variety of other purposes. “Demand,” then, will be used as the entire *schedule* of amounts a buyer will take at various prices. And the demand that has significance for the determination of prices is demand expressed in money terms in the market.✓ The fact that most people would like to have more nice clothes than they have is of significance in the market only to the extent that they can and will pay for the additional clothes.

The weekly demand of a hypothetical consumer for some commodity, say eggs, may be represented by a simple schedule, such as is given in Table 3—1. The schedule indicates that at a price of 60 cents per dozen,

TABLE 3—1

INDIVIDUAL DEMAND SCHEDULE FOR EGGS

<i>Price in cents per dozen eggs</i>	<i>Dozens of eggs purchased per week</i>
60	0
50	1
40	2
30	3
20	3

this consumer would not buy any eggs at all, at a price of 50 cents he would buy 1 dozen a week, at a price of 40 cents 2 dozen, and so on. At lower prices, eggs may be cheap enough to use in salad dressing and in angel food cakes, in addition to being used for other purposes. For this particular consumer it appears that a reduction of price generally stimulates consumption, but that a reduction from 30 to 20 cents will not

increase consumption any further. Whether a drop to 10 cents would make any further difference we do not know, since the schedule covers the range only through 20 cents.

Of course this hypothetical demand schedule is very much simplified. It might have been made more detailed by putting the price changes in jumps of 1 cent instead of in jumps of 10 cents, and by showing adjustments in terms of changes of less than a dozen eggs at a time. It would also have been more accurate to specify the grade of eggs involved, since different grades of eggs are in some respects different products, for which people have different demands.

Aggregate demand

The storekeeper is seldom particularly interested in any one individual's demand for any given product. Instead he is interested in the total demand of his customers for a product. Similarly, the economist is generally interested in aggregate (market) demand, rather than in the demand of any one individual. Aggregate demand is simply the summation of all the individual demands in any particular market area. To extend this study of the hypothetical demand for eggs, we might suppose that the individual represented in Table 3—1, whom we may call A, is only one of five people in this particular market for eggs. A's demand is shown in Table 3—2 along with the demands of the other four

TABLE 3—2
AGGREGATE DEMAND SCHEDULE FOR EGGS

<i>Prices in cents per dozen</i>	<i>Dozens of eggs taken per week by:</i>					
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>Aggregate</i>
60	0	1	0	2	0	3
50	1	2	0	2	1	6
40	2	2	0	2	1	7
30	3	3	1	3	1	11
20	3	4	1	3	2	13

consumers, and the aggregate demand is simply the summation of these five individual demands. Thus, at 60 cents 3 dozen will be taken by the entire group of consumers, at 50 cents 6 dozen and so on. Aggregate (or market) demand is the series of amounts which will be bought at different possible prices in the given market area per week or per month or per year. It is an entire schedule, a series of different amounts which will be bought at different prices, just as is any individual demand. It makes no more sense to say that the aggregate demand for a good is some absolute amount than to make such a statement about an individual's demand.

Graphic Representation of Demand

Demand may easily be shown in graphical terms, and although graphical presentation tells nothing more than does a tabular schedule, graphs

are often useful short cuts toward understanding economic processes. We may graph individual demand schedules, and similarly we may graph aggregate demand schedules.

First, let us take the hypothetical individual demand schedule of Table 3—1 as an illustrative case. This is represented graphically in Figure 3—1. In accordance with an almost universal custom of American economists, the graph shows quantity (dozens of eggs) along the horizontal base line and dollars and cents (price per dozen) along the vertical axis. The demand curve is plotted from the demand schedule given in Table 3—1 in the following manner: At 20 cents per dozen the consumer will buy 3 dozen eggs per week; therefore we place a dot opposite 20 cents and directly above the 3-dozen mark, showing that at this price 3 dozen

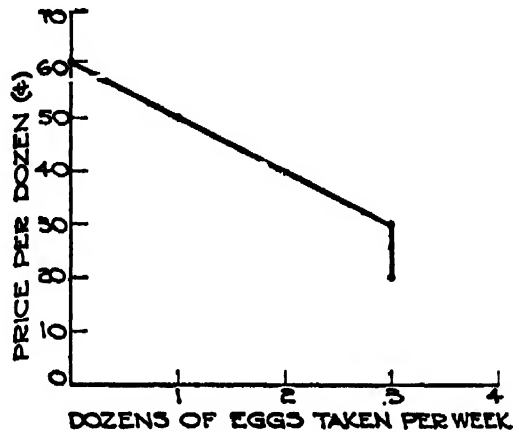


Fig. 3—1. Individual demand curve for eggs.

eggs will be bought. At 30 cents per dozen 3 dozen eggs will still be bought; therefore we place another dot, this one opposite 30 cents and still directly above the 3-dozen mark. At 40 cents per dozen, only 2 dozen will be bought; therefore we place a dot opposite 40 cents and directly above the 2-dozen mark. A similar procedure is followed for the 50-cent price. For the 60-cent price the dot is at zero, since our consumer will buy no eggs at that price. If we had taken smaller price changes, say 1 cent at a jump instead of 10 cents, the dots would be very close together and would appear to form a solid line. Although the points in our graph (Figure 3—1) are much further apart, they are joined together to show a continuous "demand curve" for eggs by our consumer. Strictly this curve may not be accurate, since we know only the points we have plotted. The stretches of the curve in between the plotted points are only approximations, and it would be only when the known points were infinitely close together that the curve would be a continuous line. However, we generally draw demand curves as solid lines, recognizing the always-present reservation that only at the actually plotted points is it a specific statement of what will be bought at those prices.

Suppose that we had come upon this demand curve as a representation of an individual's demand for eggs, instead of having plotted it ourselves. What would it tell us? It would tell us how many dozen eggs our hypothetical egg-buyer would buy per week at various prices. To find out how many he would buy at 30 cents per dozen, we could simply look at the 30-cent price and see that at that price 3 dozen is the quantity indicated. Therefore, we would know that he will buy 3 dozen at that price. We could carry out the same procedure for any other price. Or alternatively, if we were selling eggs, we could ask what price can we get from this individual if we want to sell him 2 dozen eggs a week? The answer would be 40 cents. The demand curve on the graph tells us just as much as does the demand schedule from which it is plotted, no more and no less. In this case both are purely hypothetical. But for



Fig. 3—2. Individual and market demand curve.

every person's demand for everything purchased, there is a real demand curve which has some particular shape. Different real demand curves vary greatly in shape, both as between different people and as between different products.

The aggregate demand for eggs, given in Table 3—2, may be plotted graphically just as was the individual demand. Figure 3—2 shows both the aggregate demand (DD) and the demand of individual A (curve dd) for eggs, on the same graph. The scale of quantities on the horizontal axis of the graph is somewhat more condensed than in Figure 3—1, merely as a matter of convenience in drawing the chart to fit the page space. It should be clear that this change in scale, though it changes the appearance of the individual demand curve, makes no difference whatsoever in the demand data shown by the curve. The same demand can have any slope when plotted as a curve, depending upon the scales adopted for the price and quantity axes.

Changes in Demand

Demand is a schedule of amounts which will be taken at different prices, other things assumed as constant. A change in demand is a }

change in the schedule. A change in demand, according to our usage of the term, does not occur when a buyer increases his purchases merely because of a decrease in price; this would be simply a movement from one point on the buyer's existing demand schedule to another point. A change in demand occurs only when the buyer will buy more or less than before at any given price.)

Graphic presentation of a change in demand

This distinction between a change in demand and simply a change in the amount purchased in response to a price change can be illustrated most clearly by use of demand curves. Suppose we draw a hypothetical demand curve of some individual for sirloin steak. This demand curve is shown as the solid line, labelled dd , in Figure 3—3. At 30 cents per pound he will buy 8 pounds per month; at 35 cents per pound he will buy

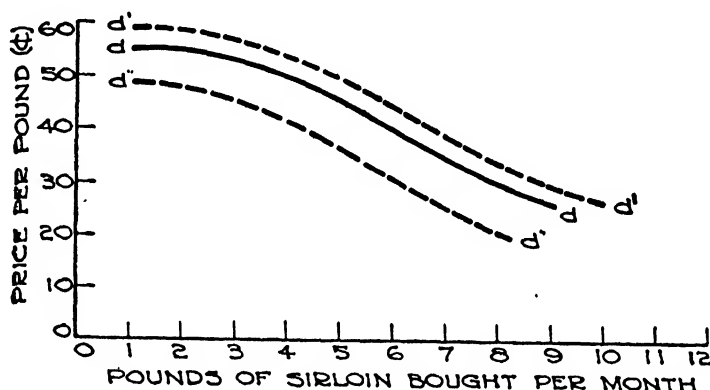


Fig. 3—3. Change in individual demand curves.

7 pounds; and so on. At any price higher than 55 cents per pound he will simply use other meats entirely and will buy no sirloin at all. The fact that he buys more or less sirloin as its price goes up or down does not indicate a change in demand.

Now let us suppose that something happens which changes this individual's demand for sirloin. Perhaps his income becomes larger and as a result his demand increases; that would mean that at any given price he would take more pounds of sirloin than before. The curve $d'd'$ represents an increase in demand over the curve dd . At 30 cents a pound, he will now buy 9 pounds per month instead of 8; at 50 cents a pound he will buy 5 pounds instead of 4; and so on. The increase in amounts taken may not be the same at some prices as at others; but whenever more will be purchased at a given price than previously there has been an increase in demand. Since quantities are measured along the horizontal axis, an increase in demand is shown graphically by the construction of a new demand curve to the right of the original one.

If the individual's demand for sirloin had decreased instead of in-

creased, the new demand curve would be to the left of the old one. The curve $d''d''$ represents such a decrease in demand as compared with dd . At each price less will now be taken than before.

Changes in individual demands lead to changes in aggregate demands. An increase in aggregate demand will be graphically represented by a new demand curve to the right of the first one, just as was an increase in individual demand. Similarly, a decrease in aggregate demand will be graphically represented by a new demand curve to the left.

Factors causing a change in demand

(A change in demand will be caused by a change in any of the factors which determine demand. Thus a change in demand may result from change in (1) income, (2) scales of preferences, and (3) availability of other goods. We shall consider each of these factors in turn.)

1. The effects of income changes on demand are obviously important. If a man's income goes up or down, he will probably purchase different quantities of eggs at prices of 60 and 50 and 40 cents a dozen than he did before the change. In short, his demand for eggs will change. So also will his demand for sirloin steak. An increase in income may lead him to buy more of both eggs and sirloin at each price. On the other hand, it may lead him to shift some of his purchases from eggs to sirloin. A good is sometimes called an "inferior" good when an increase in income leads to a decrease in demand for that good; graphically this would be represented by a shift of the individual demand curve to the left. If many individuals regard a good as "inferior," the aggregate demand for that good may decrease when incomes are rising. This might happen to the demand for low quality textiles, or for oleomargarine. Most goods are, however, "superior" goods to such a large proportion of the population that increased money incomes will lead to increased aggregate demands. Conversely, decreased money incomes will usually lead to decreased aggregate demands for most goods.

2. A shift in the ranking of items in an individual's scale of preferences will also lead to changes in demands for the goods involved. If for some reason an individual comes to like sirloin better than other meats (or comes to prefer it to some other things on which he spends his income), his demand for sirloin will increase even without any change in income. Graphically this would be shown in the construction of a new individual demand curve for sirloin, a curve lying to the right of the first one. If for some reason the individual's preferences shifted away from sirloin, the result would be a decrease in demand (a new curve lying to the left).

3. Any change in the availability of other goods may change an individual's demand for steak or eggs. For example, suppose that fresh sea foods were made more easily available in the Middle West because of improved facilities for shipping and refrigeration. It may be that shrimp and oysters have always ranked high on a consumer's scale of preferences but since they were previously unavailable the individual purchased

sirloin instead. Now that he can get oysters he will substitute some oysters for some of the sirloin he would previously have consumed at each price of sirloin. His demand for sirloin would be decreased. Graphically this would be illustrated by a new demand curve to the left of the first one. If, on the other hand, some good previously available is withdrawn from the market, the demands for its substitutes would increase.

Or to take another example, suppose good recordings of symphonies and operas to become more available. Such music may have ranked high on an individual's scale of preferences but since it was previously unavailable he had played his victrola only infrequently and used very few victrola needles. Now that he can get more and better victrola records his demand for victrola *needles* will increase, being represented by a new demand curve drawn to the right of the initial curve. The changed availability of other goods, in this case victrola records, has led to a change in his demand for victrola needles.

Availability includes price as well as physical availability. If oysters and shrimp decrease in price a consumer with given income and scale of preferences will probably buy more of these foods than before, shifting to a new position on a given demand schedule for oysters. This is not a change in the demand for oysters. But it may involve a change in demand for sirloin. Since oysters now cost less than before, some consumers will substitute them for sirloin; less sirloin will be taken at each price of sirloin, and the demand for sirloin will have decreased. If, on the other hand, the price of oysters had increased, consumers might increase their demands for sirloin. Thus a change in the price of a good leads to (1) a new position on a given demand schedule for that good, and usually (2) changes in demands for some other goods.

The Downward Slope of Demand Curves

Perhaps the most noticeable thing about these demand curves is that they slope downward from left to right; that is, they say that other things being equal the lower the price the more that people will buy. This statement certainly seems nothing but a common-sense observation; yet it bears careful examination. We shall look first at factors determining the slopes of individual demand curves; then we shall consider aggregate demands.

The downward slopes of individual demand curves

Any drop in the price of eggs, all other things remaining unchanged, will make them a "better buy" relative to other things than they were before. An individual with a given income and scale of preferences will find that for any given amount of eggs he will make less sacrifice of other things than before. By substituting more eggs in his diet for some of the meat and cheese he previously consumed, he may save enough to get something else that is important to him but that was previously beyond his reach. Moreover, he may substitute eggs not only for meats and cheeses, but for some other commodities that seem to have very little relation to eggs.

Many an economic policy, whether of government, business, or private individual, has gone astray for lack of attention to elasticity of demand. The corner grocer ordering stock for his shelves must know how much of each single item he can move at the price he has been charging. He must have an eye to the demand for No. 2 cans of Libby's Tomatoes, No. 5 cans of Dole Pineapple Juice, and some hundreds of other items. But for the smart grocer this is not enough. If he cuts his price on one item, say from three cents to two, he will expect to sell more, but he wants to know how much more. If sales increase greatly his gross revenue will be more than before, and it may be that his profits will be greater, but if sales increase only slightly his gross revenue may drop and certainly profits will be less. What he wants to know then is the elasticity of demand for each individual item he sells. On those items where demand is elastic it will pay him to charge relatively low prices, while on those whose demand is inelastic he can do better with a higher price.

Some historical examples also serve to illustrate the importance of the elasticity principle. One can raise gross income by raising prices *only* when demand is inelastic. Failing to recognize that demand was elastic, the Eastern railroads in the 1930s tried to increase passenger revenue by raising fares. But travelers responded by traveling much less, or shifting to other means of transportation. In the end it was found that reduction rather than increase of fares was the way to increase revenues. More than once, too, have great public utility corporations fought in the courts against proposed rate reductions, only to find to their embarrassment after losing the case that the new rates were more profitable than the old. It seems such a simple and obvious truth that higher prices bring in more money; yet many an executive has learned about demand elasticity the hard way. The same error has occurred among labor unions when they have proceeded on the assumption that higher wage rates necessarily mean higher incomes.

Factors affecting demand elasticity

What causes the demand for any particular good or service to be elastic or inelastic over any given price range? Under what conditions, in other words, would we expect buyers to be highly responsive to price changes, and under what conditions not?

The elasticity of aggregate demand depends both on the elasticities of individual demands and on the appearance of new buyers or disappearance of old ones with changes in prices. (1) Influences leading toward inelastic individual demands, (2) influences leading toward elastic individual demands, and (3) the elasticity of aggregate demands must therefore be considered separately.

1. Under certain circumstances, the demand of an individual consumer is likely to be *inelastic*. Would you be likely to change markedly your consumption of matches, or salt, when their prices changed by one fifth? Probably not; probably your demand for these things is inelastic. Why? Some of the reasons may be classified as follows:

a. Your expenditure is probably extremely low. Both matches and salt make up a very small part of your total budget. Even a very substantial percentage increase in the price of either would fail to make it a very significant expenditure item. If the price of matches fluctuated between 50 and 60 cents per box instead of 5 and 6 cents, you might pay more attention to a change of price.

b. Your want may be urgent. Both salt and matches are goods which you probably consider essential. If they rise in price you would probably rather give up something else than forfeit salt on your food or matches to light your cigarets, if such a choice became necessary. What things you consider "necessary" depends in large part on the availability of substitutes. If there were a good substitute for salt, your want for it would be much less urgent and your demand for it much more elastic. Urgency of want depends also in large part on habit and social custom. Your demand for cigarets may be highly inelastic because you have fallen into the habit of smoking. Whatever the underlying causes, urgency of want tends to bring inelastic demand.

c. Your demand may be a joint demand. Salt and matches are demanded jointly with some other good or goods. Salt is demanded jointly with most food. Since the major portion of the total expenditure is on the foodstuffs, it is unlikely that an increase in the price of salt would make much difference in the amount of it you would buy. To give up the salt would virtually mean giving up eating the food. Matches are demanded jointly with cigarets, and are merely a small part of the cost of smoking.

2. By contrast, your demand for certain articles, such as football tickets, shoes, butter, and branded breakfast foods, is likely to be fairly elastic. Why? In part the reasons are the obverse of those leading to inelastic demand. Factors leading to elastic demand may be classed as follows:

a. Your expenditure may be a costly one. The higher the portion of your total budget involved in any expenditure, the more strongly you are likely to respond to changes in the price of the article. Football tickets are in a price range high enough to make significant dents in the incomes of most college students. Demands for these tickets are therefore sometimes highly elastic.

b. Your demand may not be urgent. The less urgent your want for any article, the more likely you are to give it up in case its price rises. Again the question of "urgency" of want is largely one of the availability of substitutes, and of habit and social customs. For example, your demand for a particular brand of corn flakes is likely to be fairly elastic since there are so many close substitutes that this particular breakfast food is not very "essential."

c. The commodity may be durable or repairable. For example, shoes are relatively durable and can be repaired for substantially less than the price of new shoes. Therefore, if the price of shoes rises very much you are likely to wear your old pair longer, and possibly to have them repaired

instead of buying new ones. The more durable and repairable a good, the higher its elasticity of demand is likely to be.

d. The commodity may have multiple uses. The more uses a good has, the higher is likely to be its elasticity of demand. Butter, for example, is considered a luxury to some even as a spread for bread. To others it may be a "necessity" on bread but a "luxury" on vegetables or as a fat for frying. These individuals will have an elastic demand for butter as its price falls since they will then feel able to use it for the "less essential" purposes. The relation between versatility of a good and the availability of close substitutes for at least some of its uses should also be noted. Thus the more versatile a good the more elastic demand is likely to be (especially in the lower price ranges), for two reasons—because it will have a lot of uses of secondary importance and because substitute goods may be readily obtained for some of these uses.

3. Since an aggregate demand schedule is the sum of individual schedules it would appear that we could simply generalize the comments of the previous paragraphs to describe factors underlying varying elasticities for aggregate demand schedules. This is true, but an additional consideration exists, which must be included. The attitude of an individual toward goods as "necessities" or "luxuries" will depend in large part on his income. Those whose incomes are high will consider as "necessities" items which are "luxuries" to the poor. For most consumers' goods the demand of the poor is probably more elastic than that of the rich. At low prices the proportion of the aggregate demand which comes from the poor is larger than at high prices. Consequently, insofar as this factor operates, the aggregate demand for an article at low prices will be more elastic than at high prices.

Interrelations of Demands for Different Things

In the broadest sense all goods and services compete with one another for the consumers' incomes. But if we are dealing primarily with a relatively small number of goods and services, it is possible to distinguish two basic types of relation—substitutive and complementary. Two goods are substitutive if an increase in the consumption of one is associated with a decreased demand for the other. Goods which are substitutes compete with one another directly. On the other hand, two goods are complementary if an increase in the consumption of one is associated with an increased demand for the other. When two goods are complementary we often say that there is a "joint" demand for them.

Cases of each type of relationship are easy to think of. Oranges and grapefruit are primarily substitutive; if you have one for breakfast you will not have the other. Bread and butter are primarily complementary; if you buy more bread you will probably buy butter to go with it. Some goods are close substitutes; some are only very imperfectly substitutable. Some goods are very close complements; some are only remotely complementary. Some goods are substitutes in some respects, complementary in

others; for example ham and eggs to some extent compete with each other and to some extent they are demanded jointly.

To indicate some of the implications of these relations we may take a very simple example of streetcar rides, movies, and victrola records. Movies and victrola records may be substitutive to some extent. Streetcar rides to get to and from the movies are complementary with movies. Suppose that for some reason carfares are raised. This will, other things being equal, decrease the demand for movies, since now the cost of movies plus carfare has increased even though movie prices remain unchanged. But since victrola records are substitutive with movies the higher price for carfares may lead to an increase in the demand for victrola records. The change in the carfare has affected the demand for victrola records with which streetcar rides have no direct relation, as well as the demand for movies which is complementary with that for streetcar rides. Such ramifications in the relations between commodities extend in a broad complex web over our entire pattern of consumption.

Producer Response to Consumer Demands

The principal motivation to production in a private enterprise exchange economy is the anticipation of money incomes (which are then spent on consumers' goods). In the effort to maximize their incomes workers seek the jobs where pay will be highest, and businessmen seek opportunities to maximize profits through the production and sale of goods and services. The opportunity to obtain an income depends primarily on the willingness of other people to pay for the goods and services the particular individual or business may offer for sale. If the demand for a product is great there will be many opportunities to make money by producing it; if that demand is small the income opportunities will be less. It is primarily in response to consumer demands directly, or to these demands as reflected in demands for the machines and raw materials to make the products wanted by consumers, that businessmen decide what their activities will be. In some cases the response is very direct. In some cases the individual businessman may be in a position to decide whether he will sell a small amount at a high price or a larger amount at a lower price; in other cases he may simply have to take market price as given and decide how much it will pay him to produce. In still other cases demand itself may be markedly affected by the activities of the very producers who are seeking to sell their products; indeed, if a man can both create a demand for a product and produce the product, he may find great profit opportunities. Even in these cases, however, there are limits on the abilities of individual producers to persuade consumers into changing their selections and their basic preferences. For the most part productive activity takes place in response to estimates of these preferences as they will be expressed in consumer purchases.

CHAPTER 4

The Unit of Business Enterprise

THE business enterprise is the unit within which productive activities are organized to turn out goods and services, for sale ultimately to consumers. Such units vary all the way from a small farm or shoe shop or cigar store to a large plantation or a big industrial corporation. Before proceeding further with a broad analysis of economic processes, it is therefore useful to examine briefly the forms some of these units of production organization may take.

The Single Proprietorship

Legal form

The oldest form of business organization is the single proprietorship, and it is still by far the most common numerically. Simply stated, a single proprietorship is a one-man business. The one man (proprietor) is not necessarily the only worker, but he is the sole owner of the business and the sole controller of its policies. He may hire any number of employees and may delegate various duties to them, but he alone is ultimately responsible for management. He may borrow funds and contract to pay interest on them. But he receives all the net profits made by the business; he is also the first to bear the losses and is responsible for all the debts.

Legally, the proprietor and his business are essentially one. All of the proprietor's personal wealth and income (with minor exceptions such as a homestead, in some states) may be taken, if necessary, to pay the debts of the business. We say the proprietor has "unlimited liability." And any assets of the business may be seized to pay his personal debts. Any competent businessman will keep separate his personal accounts and those of his business so that he may know accurately whether or not the business is making a profit; but the law takes no cognizance of such separation of accounts. In view of this situation, a proprietor is naturally hesitant to delegate authority to anyone else in important questions of management and policy.

Advantages

The advantages of the single proprietorship as a unit of business enterprise can be quickly summarized. The proprietor is his own boss, he

need take orders from no one—a consideration of major importance for many persons. In a small business, management by a single proprietor is likely to be more efficient than would be division of authority between two or more managers. Being solely responsible for the management of his business, the proprietor need not worry about the danger of unwise decisions by associates. (The fact that he has sole responsibility may also lead him to be especially careful in making decisions.) And a single proprietorship can be started and carried on with a minimum of legal formalities and expense; it is free from special regulations and taxes imposed on corporations.

Disadvantages

In spite of this list of advantages, the single proprietorship has a major drawback which has led to its replacement by other forms of enterprise in large areas of economic life. It is very difficult, if not impossible, under a proprietorship to raise huge sums of capital for carrying on the large-scale production which has become so necessary in producing many products today. For a retail store, or a farm, a fairly small sum suffices to set up business on an efficient scale. But for an automobile factory or a steel mill, a very large investment is required to set up and operate a plant which can obtain the economies of large-scale production necessary to compete with other large-scale producers. There are very few individuals with personal fortunes sufficiently large to finance such undertakings alone. And even when an individual does have a fortune large enough to permit his financing a huge business enterprise, he is probably hesitant to do so (1) because of the liability of his personal fortune for the debts of the business, should the business prove unsuccessful, and (2) because of the difficulty of liquidating an investment so made, if he wishes to sell out, or if he dies (in which case the life of the business is legally terminated). Still another disadvantage is the inability of any proprietor to delegate authority to subordinates without endangering his personal fortune, and such delegation of authority is necessary for large-scale business enterprise.

Decreasing importance

In view of these drawbacks, individual proprietorships have steadily become less important in American economic life. They are of little importance in manufacturing, transportation, communications, mining, construction, and the utilities, which together account for roughly two thirds of the total employment and of income paid out in our economic system. But, in spite of their drawbacks, individual proprietorships have remained the prevalent form of business organization in agriculture, in retailing, and in service enterprises (doctors, lawyers, hotel operators, and so on). There are still some nine million individual proprietorships, almost seven million of them in agriculture, including tenant farmers as well as owner-operators. In 1936 only about 8 per cent of all farms were

operated by partnerships or corporations, although only 57 per cent of the individual agricultural proprietors owned their own farms. Of the other two million individual proprietorships, over half a million were in retail and wholesale selling and over half a million in service enterprises. This continued importance of individual proprietorships in agriculture, retailing, and services is explainable largely on the ground that these are the outstanding cases where small-scale enterprise is at least as efficient as large, and often more so. Even in these fields there is some evidence that individual proprietorships are gradually becoming less prevalent. Especially is this true in retailing, where such large-scale enterprises as mail-order houses, chains, and department stores have been expanding rapidly in recent years, replacing small-scale individual proprietorships. There seems little likelihood, however, that the individual proprietorship will soon be generally replaced as the common form of organization in farming, retailing, or services, because many of the particular tasks in these lines are well adapted to relatively small-scale production.

The Partnership

Legal form

A partnership consists of two or more persons who are joint owners and proprietors of their business. Usually there is a written partnership agreement ("Articles of Partnership") which specifies the capital and services which each is to contribute and the manner in which the profits or losses are to be shared. However, this agreement may be only an oral one. In general the law views a partnership very much as it views an individual proprietorship—the partners are legally responsible for the debts of the business to the extent of their personal fortunes. In case the business cannot meet its obligations, the courts will first attempt to collect from the individual partners in the manner specified in the Articles of Partnership, wherein their respective responsibilities may be listed, but if any partner cannot pay his share, all of the other partners become personally liable for it. For example, if there were four partners in a bankrupt business and three of them were unable to pay anything on the debts of the business, the fourth would be personally liable for all the debts. Legally the fourth could then collect from each of the other three his share as specified in the partnership agreement, but actually this is likely to be very difficult if the creditors of the business have already been unable to collect.

Furthermore, each partner is liable for all the acts of other partners taken in the name of the business, in spite of any private agreement the partners may have among themselves. Thus, the partners may have definitely agreed that one only is to make sales, another only to manage the plant, another only to purchase materials, and so on; but if one of them foolishly incurs an unwise obligation for the firm, each other partner is completely liable for this obligation if the business cannot meet it. Again, the other partners might sue to collect from the erring member

if he had broken an agreement, but whether or not they collected they would each be liable for the debt of the business.

The reason for this strict legal attitude is essentially this: Any person dealing with the partnership cannot be expected to know all the details of the partnership agreement. When he deals with a partner he has a right to assume that he is dealing with an authorized agent of the business; to force a buyer or seller to investigate the articles of partnership before dealing with any partner to be sure of that partner's right to make such a contract would be unjustifiably awkward and burdensome. Therefore the law protects the outsider by assuring that the business will be liable for the contracts made by any partner.

"Limited partnerships" are permitted in some states. A limited partner is one who is liable for the debts of the business only up to some specified figures, say \$10,000. Such a partner, according to most state laws, may contribute capital to the business but not his services, and his name may not appear in the partnership's name. Moreover, he has no right to do business in the partnership's name. This provision permits partnerships to obtain capital from wealthy individuals who are willing to make some investment but are unwilling to make themselves completely liable for the operation and obligations of the business. Such partners are sometimes called "silent" or "sleeping" partners. But even where a limited partnership is permitted, the partnership must have at least one, and usually two, general partners who are liable for the debts of the business in the manner indicated above.

Advantages

A partnership has one substantial advantage over an individual proprietorship. By including more proprietors, the partnership is better able to obtain more capital for setting up and operating the business. This is so (1) because of the larger personal resources on which the business can draw, and (2) because of the stronger credit position of the business in seeking to obtain borrowed funds when its debts are backed by the personal resources of several partners. Another advantage of the partnership is its ability to obtain partners each well adapted to some phase of the business' operations, thus having a proprietor in charge of each phase rather than simply a hired hand. As contrasted with the corporation and other large-scale forms of organization, the partnership, like the single proprietorship, retains a more "personal" character, which may be very important in retailing and direct service enterprises. Furthermore, the partnership is free from special corporation taxes and much of the red tape and special regulations imposed on corporations.

Disadvantages

Although it is superior to the single proprietorship in raising capital, the partnership is still likely to have difficulty in raising very large sums. In the first place, there are few partnership groups wealthy enough to put up the funds necessary for present day large-scale business operations.

And even when there are wealthy individuals, they hesitate to expose themselves to complete liability for the debts of the business and the acts of their partners. There are, of course, important exceptions, but in general the partnership has been unable to provide the capital needed for large scale enterprise.¹ There are other drawbacks to the partnership. If the partners disagree on major matters of business policy, there is no solution short of dissolution. The life of the partnership is limited to the lives of its partners: the death of one partner legally dissolves the partnership, as does the withdrawal of a partner for any reason. A partner wishing to sell out his share can do so only if all other partners are willing to accept the purchaser. Otherwise, the alternatives are for the remaining partners to buy out the retiring one, or for the partnership to be dissolved.²

Extent of the partnership form of enterprise

The partnership as well as the single proprietorship is subject to the fundamental drawback of difficulty in raising capital. Therefore, as might be expected, although there have been some large partnerships, the partnership has never become a prevalent form of business organization. In 1936, there were about 230,000 partnerships in the United States. Of these, about one third were in retail or wholesale selling, about 15 per cent in services (professions, hotel operators, and so on), 13 per cent in financial enterprises, and the rest scattered. In spite of the few well-known large partnerships, the average size, measured by employment, gross income, and net income, is relatively small.

The Corporation

Legal form

A corporation is a business organization which exists as a "legal person" apart from the individuals who own and control it. A corporation may carry on business in its own name; enter into contracts, sue and be sued, own property, borrow and loan money. In general it may as a business unit do all the things which any individual person may legally do in business. A corporation may have one or a number of purposes. Many of the earliest corporations in England, where charters were derived from the crown, were religious, fraternal, and educational institutions. For a long time, corporation charters specified rather closely what the purpose of the corporation was to be and limited the activities of the corporation to achievement of this purpose. Today in this country the corporation charters granted by most states are, however, so broad in their grants of power and so vague in their statements of corporate purposes that they

¹ Until recently J. P. Morgan and Co. was an example.

² The legal provisions regarding partnerships vary somewhat from state to state. The statements made here apply in most states, and in general the differences are those of detail.

exercise little effective restraint over corporate activities. Indeed, since corporations pay fees for charters and special taxes to states within which they are chartered, some states have competitively made corporation charters more and more lax in an attempt to get more fees. Delaware has gone furthest in this competition for corporation charters, and Wilmington contains the "dummy offices" of hundreds of large business firms whose real offices are in New York and Chicago.³ It should be noted that no matter in what state a corporation takes out its charter, it is permitted by the interstate commerce clause of the federal constitution to carry on business in any other state. Thus it is virtually impossible for any state to regulate corporation practices effectively, for any attempt to do so is likely simply to result in the corporation legally moving to a more lax state, even though it carries on the same business in the same places as before.

The financing of corporate enterprise

The corporation has numerous advantages as a form of business organization, but much the greatest of these is its superior ability to raise large amounts of capital. To understand this major advantage of the corporate form, and later to understand some of the abuses made of it, it is necessary to know something of the ways in which a corporation gets its funds.

A corporation is financed by the investment of a few or many individuals in "corporate securities." These individuals may simply make loans, or they may be part owners of the corporation. Corporate securities thus fall into two broad classes: stocks, which represent ownership in the corporation, and bonds, which represent money loaned to the corporation by bondholders. There are many variations within each of these two categories, and the distinction between some stocks and some bonds may be very slight and therefore should not be overemphasized. The most important characteristics of corporate securities as forms of investment are their variations with regard to: (1) the right to an early claim on the income of the enterprise before other security holders receive any payments, (2) the right to the "residual" income, however large, after others have been paid promised amounts, and (3) the right to vote on personnel and policy in the corporation, and hence the power to control the corporation.

Common stockholders are owners of a corporation. Common stock is usually "voting stock" and the stockholders have the right to elect the board of directors (except under special circumstances); the stockholders are therefore in a position to control the policies of the corporation so long as it remains solvent. They are entitled to any income remaining after prior claims of creditors and other owner-investors have been met. If the corporation is dissolved, they are entitled to all that remains (if

³ Delaware granted the United States Steel Corporation a charter permitting it "to engage in any kind of business in any part of the world."

anything) after everyone else has been satisfied. The common stockholders are the "residual claimants" to the corporation's income and property. They gain the most when income is high, and they are the first to lose when things are not going so well. Net income earned and paid out to stockholders is called dividends. Although the net income of the business "belongs" to the stockholders, often the corporation does not pay it all out but instead reinvests a part in the business. When this is done we often say that the corporation has "plowed back earnings"; such reinvested funds are called "surplus." Since the stockholders have the residual claim on the assets of the corporation, increasing the assets by reinvesting net income increases the value of the stock. Whether the net income is paid out or reinvested, it accrues to the stockholders.

In some cases, common stock has been classified into two types "non-voting" and "voting" (usually called "A" and "B"). The "B" stock carries voting power in electing directors, but the "A" stock does not. An example is American Tobacco A and American Tobacco B, of the American Tobacco Company (makers of Lucky Strikes). Aside from this difference in voting power, the two types of stockholders ordinarily take equal risks and share the opportunity to get large dividends. The issue of nonvoting common stock makes minority control by "insiders" possible with only a small investment while yet preserving flexibility in the concern by maintaining a high proportion of "owner-financing" rather than by increasing bond issues; frequently nonvoting common stock is issued to the employees of a concern in "profit-sharing" schemes.

There are a number of important differences between the position of the stockholder owner of the corporation and that of the owners of a partnership: (1) The owners of stock are not liable for the debts of the corporation beyond the amount they pay for the common stock when they purchase it. If the corporation goes bankrupt, creditors can only seize the assets of the business; they cannot call on the stockholders to make good on the debts of the corporation.⁴ (2) The stockholder as an individual has no power to act in the name of the corporation. It is therefore usually a matter of no importance who the owner of the stock is unless a particular group of men is seeking to gain control of the corporation. The actual management of the corporation is delegated to the board of directors by the vote of the stockholders, and by that board to the officials it selects. (3) Investment in stock is a "liquid" investment as compared with investment in a partnership. The individual investor can withdraw by selling his stock, and new investors can join in ownership of the corporation by the purchase of stock.

⁴ There are minor exceptions. In many states, stockholders of insolvent banks have been liable for assessment of 100 per cent of the face value of the stock, in addition to the amount originally paid for the stock. This provision was made to give special protection to depositors by making more sure they would receive their deposits back even when banks failed. This provision has, however, been eliminated in most states and commonly bank stock liability is now on a par with other stocks.

Bondholders are creditors of the corporation; stockholders are owners of the corporation. Bonds represent money loaned to the corporation. A corporation might borrow money from the bank by signing a note promising to repay the money, with interest, at some future date. But when corporations want to borrow large sums for long periods, they commonly do so by issuing bonds which are sold to persons or institutions having funds to invest. Bonds are nothing but promises of the corporation to repay to bondholders the funds borrowed, at some specified future date with a set rate of interest over the period during which the bond is outstanding.

Bondholders, being merely creditors, ordinarily have no voting power to elect directors, and hence no direct control over the policies of the corporation. They take much less risk than the stockholders, since a specified interest on bonds must be met before any income is available for dividends to stockholders.⁵ On the other hand, most bondholders receive only the set rate of interest, even though the corporation makes huge profits. If the corporation should go bankrupt and be liquidated, bondholders would be paid before any payments were made to stockholders, since the bonds represent debts of the corporation which must be paid before anything is paid to the owners; but the bondholders receive back only the funds loaned. Many bonds contain provisions that if a corporation fails to pay interest when due, voting power is automatically given to bondholders so that they can protect their interests by having a voice in directing the policies of the business. In some cases, if interest is defaulted several times running, stockholders may lose the right to vote and complete control may be transferred to bondholders until past interest defaults are cleared up.

There are numerous types of bonds, and large corporations typically issue several types. Some are mortgage bonds, backed by specific pieces of physical property; some are backed by specific holdings of the corporation in securities of other corporations; some are backed just by the general assets of the corporation, and so on. The important factor in determining the degree of risk or security of a particular issue of bonds in a given corporation is not whether or not the bond has some specific backing, but whether it has first, second, third claim on the income and property of the corporation. In each corporation there is a definite rank of priority of claims on earnings, with first-mortgage bondholders ordinarily having "first lien," that is, first claim to payment. In fact, securities may be ranged in order of priority of claims on earnings and assets in a continuous series from first lien bonds to common stocks.

It is easy to overemphasize the formal distinctions between various types of stocks and bonds and the formal rights of the various security holders. This is easily illustrated. When a corporation becomes officially bankrupt and unable to meet interest or principal payments to bondholders, the corporation may be liquidated and payments made to

⁵ "Income bonds" are an exception.

security holders in the order indicated. Usually, however, the corporation is worth more as a going concern. A compromise agreement is therefore worked out whereby all classes of security holders make some sacrifice, scaling down their claims in interest payments in order to keep the corporation in operation. In such a compromise ("reorganization") even first lien bondholders may have to agree to take a lower rate of interest, although they seldom have to make much of a sacrifice. On the other hand common stockholders may be completely wiped out, with (let us say) third and fourth lien bondholders giving up their bonds and taking stock in the reorganized corporation. In general, the order of priority holds as to sacrifices made, but the formal distinctions are often blurred. Such a reorganization may have little effect on the actual assets and the services rendered by the corporation, and may change only the financial structure of the corporation, usually scaling down the annual interest charges.*

Preferred stockholders have a position intermediate between that of the common stockholder and the bondholder. Preferred stock sometimes carries a vote, though very often not. It typically has a set rate of dividends, say \$6 per share, that must be paid it before any dividends can be paid on the common stock. Preferred stock usually has preferment if the corporation is liquidated (that is, preferred stockholders will usually receive priority over common stockholders in the distribution of assets). On the other hand, preferred stock is definitely inferior to bonds in priority of claim on income and assets. If the preferred stock is *cumulative*, dividends not paid during bad years must be paid up on this stock before any dividends can be paid in subsequent years to the common stockholders. If the preferred stock is *noncumulative* and the preferred dividend goes unpaid any year, preferred stockholders have no claim for these missed dividends in future years when profits are being made. If earnings are very large, the preferred stock may or may not "participate" in these exceptionally large earnings. If the preferred stock is *participating*, after the common stock has received some specified return per share, common and preferred will share the remaining exceptional earnings. If the preferred is *nonparticipating*, it receives only its set dividend, regardless of how large the profits of the business may be, and all profits above preferred dividends go to the common stockholders. Although preferred stock may appear to be a happy compromise for the investor seeking low risk with the possibility of substantial rates of return, there is no type of security which is so likely to be misleading, especially if it is noncumulative and nonparticipating. The only way to tell whether preferred stock is really "preferred" in an important way is by a careful examination of the provisions for that particular preferred stock.

*Such reorganizations have been very common. For example, three fourths of the major railroads of the country have gone bankrupt at least once in the last half century, many of them two or three times.

Although *within the same corporation* bonds are a more conservative investment than stocks, it should be noted that this statement does not necessarily hold in comparing the securities of different corporations. American Telephone and Telegraph common stock has paid \$9 per year dividends through thick and thin over the last decade and a half, paying in depression years out of surpluses accumulated in good years. On the other hand, many railroad bonds, considered very sound in good years, were defaulted completely during the depression. In evaluating different securities as investment opportunities, there is no substitute for expert knowledge and for careful investigation of the securities in question.

Advantages of the corporate form of enterprise

The advantages of the corporate form of organization as a method of financing large-scale enterprise are many. It makes possible *centralization of responsibility for management, continuity of life of the organization apart from the life of the owners, and the attraction of funds from many sources*. The techniques by which this is made possible and their significance may be summarized briefly as follows: (1) All owners of the corporation (whether holders of common or preferred stocks) incur "limited liability," the maximum possible loss being determined by the amount of the investment. Their personal fortunes are not subject to seizure in the event that the business goes bankrupt. This obviously increases the attractiveness of investment in corporation stocks. (2) The individual owners of a corporation cannot act independently in the name of the enterprise. Instead they elect a board of directors, which appoints a salaried executive manager to whom will be delegated the powers necessary for conducting the business. This centralization of management is obviously essential in a large organization with many members. It also has the effect that the personality and the credit of individual members of the corporation are matters of relative indifference to other members. Initiation and expansion of the enterprise are very much simplified, and multiplication of membership is made possible by removing management powers from the members as individuals. Since members are easily replaced, the life of the corporation is independent of the life of its members, and continuity is facilitated. Finally, associated with the divorcing of management rights from ownership is the greater marketability of investments in corporate enterprise. Stocks are readily bought and sold. This characteristic of corporate securities has the added advantage of attracting investment by savers who wish to keep their assets in a reasonably "liquid" form. (3) There is a wide range and variety in corporate securities, both stocks and bonds. This makes possible appeals to many different types of investors. Stocks are available for those primarily interested in an investment which may increase in value over time, and who are willing to take the risks associated with this opportunity. The large investor interested in gaining control over an enterprise will of course purchase stocks. In any given corporation, bonds of various sorts are more conservative investments

than are stocks of the same company, and they will attract those interested primarily in an investment which will give a fairly steady moderate income. Preferred stock gives a possibility of higher rates of return than do bonds in the same corporation, but at greater risk (even when it is "cumulative"). Since all these securities are ordinarily sold in fairly small denominations, they are available to small investors whose resources would not be so readily tapped without such securities.

Disadvantages of the corporate form of enterprise

The major disadvantages of the corporation as viewed by those interested in establishing a business are the costs of the financing operations and the special taxes and regulations imposed on corporations. When corporation charters are granted, some states impose definite regulations as to the scope of corporate action, although this drawback is not very serious, owing to the ease of incorporating in lax states. In any state the process of incorporation involves considerable expense. The listing of corporate securities for sale on a major stock exchange now also involves extensive red tape and expense, since new securities must pass inspection by the Securities and Exchange Commission before they can be listed. Moreover, the federal government and most states impose a special income tax on corporations, and in some states other special taxes are imposed. In many cases these taxes do not actually impose any differential burden on corporations, but rather aim to tax them on a basis of equality with other forms of business organization; in other cases corporations have definitely been subjected to discriminatory taxation. However important these drawbacks may have been in some cases, they have by no means been sufficient to offset the advantages of the corporation in other respects.

Prevalence of the corporation

Largely as a result of the great superiority of the corporation as a device for raising capital and centralizing control, it has become the prevalent form of organization in all fields where large-scale investment is necessary for low-cost production. Today the corporation holds virtually undisputed sway in manufacturing, mining, transportation, communications, construction, utilities, and finance. Agriculture, trade, and services are the only fields where corporations are not predominant. Table 4—1 indicates the extent to which corporations have taken over the field in manufacturing. Their predominance is clearly indicated in the first three sets of figures even though a small majority of manufacturing enterprises are still noncorporate businesses. Corporations are typically larger than other enterprises; some are Goliaths. In 1935, total assets of the American Telephone and Telegraph Company were \$3,998,000,000; those of the Pennsylvania Railroad \$2,863,000,000; and those of Standard Oil of New Jersey \$1,895,000,000—to cite leading examples. In the same year, 659,000 different persons owned American Telephone and Telegraph stock, though this is an extreme case. The

200 largest nonfinancial corporations in 1935 had total assets of \$74,232,000,000. No one of these had assets of less than \$67,000,000.

TABLE 4—1 *

MANUFACTURING CORPORATIONS IN THE UNITED STATES

Year	<i>Value of corporation output as per cent of total output</i>	<i>Workers in corporations as per cent of all workers in manufacturing</i>	<i>Value added by corporations as per cent of total value added</i>	<i>Corporations as per cent of total establishments</i>
1904	73.7	70.6	71.9	23.6
1909	79.0	75.6	77.2	25.9
1919	87.7	86.6	87.0	31.5
1929	92.1	89.9	91.5	48.3
1939	92.6	89.4	...†	51.7

* Derived from data in the U. S. Censuses of Manufactures; subsections dealing with types of organization of manufacturing establishment.

† Figures omitted because of change in classifications.

The corporation as a device for concentrating control over large investments

Perhaps the most striking characteristic of modern business enterprise is the great concentration of control which has developed in some fields since the Civil War period. Today the gigantic corporation spreading over the entire country and producing a variety of related products is a commonplace; seventy-five years ago a billion-dollar corporation would have been a wild man's dream. Aside from its potentialities as a money-raising device, the corporate form lends itself particularly well to use by minority groups in obtaining control over large investments with a minimum of contribution on their own part.

There are two principal ways in which minority groups may gain control of large investments in a corporate enterprise: (1) the use of different types of securities, and (2) a combination of inside pressures with the lethargy of the majority of stockholders in using their votes.

1. Since only a majority of the voting stock is necessary to control the policies of a corporation, it is obvious that the corporate form enables persons to control an enterprise with only slightly over half the total investment in any case. But the use of nonvoting securities makes it possible to obtain and exercise control with far less than half the total investment.

A simple hypothetical example will illustrate this possibility. Suppose a group of promoters to be organizing corporation X, whose total required invested capital amounts to \$10,000,000. The organizers issue \$5,000,000 of bonds, \$2,500,000 of nonvoting preferred stock, and \$2,500,000 of common stock. Since only the common stock has voting power under ordinary circumstances, it is necessary to control only the common to elect the board of directors and control the corporation. Half the com-

mon would amount to \$1,250,000, so an investment of slightly over this amount would be sufficient to control the ten million dollar company. In many cases the proportion of nonvoting to voting securities is larger than this, so that control may be obtained with even a smaller investment relative to the total. And in some cases even the common has been divided into "A" and "B" classes of which only the "B" has voting power.

The significance of this diversification of securities for the concentration of corporate control is obvious. At the same time it has a further broad significance in encouraging the use of "top-heavy" financial structures, in which the debt securities (bonds) are very large relative to the equities (stocks). This means that in good times the relatively small investment in stock will receive large rewards in dividends, since the bonds (and often the preferred stock) receive only a fixed amount regardless of total earnings. On the other hand, in bad times the heavy fixed debt burden is likely to prove very oppressive to the business: not only is there likely to be nothing left over for the stockholders after the heavy debt charges have been met but there is even likely to be an inability to meet these charges, which may lead to bankruptcy and complete loss of the equity of stockholders. From the point of view of mitigating business cycles, avoidance of such top-heavy debt structures is very important.

2. Having raised funds by sale of securities, corporation "insiders" often find it easy to retain control with an investment of less than half the voting stock. The policies of almost every corporation are dominated largely by those stockholders who are closely connected with the operation of the business; in many cases large numbers of stockholders know virtually nothing about the operation of the enterprise in which they hold shares of ownership. This lethargy of the "outsiders" among the stockholders is especially apparent in the very large corporations which have thousands of stockholders, many of whom hold only a few shares. To be sure, such stockholders receive notices of the annual meeting of the corporate shareholders for electing directors and transacting any other business which may arise, but receiving such a notice is a far cry from attending and actively participating in the meeting. If I own ten shares of General Motors, the chances are very strong that I have bought it purely as an investment. I know perfectly well that my ten votes would be of virtually no significance in electing any particular directors to the board and in determining the policy of General Motors. If I live in Washington and the meeting is held hundreds of miles away, the chances are very slim that I will spend the time and money involved in going to the meeting, merely to watch a routine election which has probably been determined well in advance by the large voting blocs of comparatively few "insiders." Of course, even though I do not go to the meeting, I am entitled to send a "proxy," designating a person of my choice to vote my shares for me. Typically I will receive a proxy already filled in designating some officer of the corporation to vote for me, and the chances are that if I send a proxy at all I will simply sign this one

and send it, thereby giving the "insiders" my votes to elect their choices. Of course, this lethargy on the part of many stockholders does not always exist. In some smaller corporations virtually all stockholders take an active interest in the conduct of the business. Even in large corporations sometimes conflicts arise in which sharp struggles for proxies ensue, with the control and basic policies of the corporation at stake. But for the most part, minorities in control can feel sure of their ability to retain effective control without holding even approximately half the voting stock. They are on the ground; many of the stockholders are far away and little interested.

The striking extent of such minority control is indicated by data collected relative to control of the 200 largest nonfinancial corporations in the United States in 1930.⁷ In 6 per cent of the corporations control was exercised by nearly complete stock ownership; in 5 per cent control was by majority stock ownership; in 23 per cent control was by minority stock ownership; in 21 per cent control was by legal devices enhancing the power of minority or managerial groups; and in 44 per cent control was by the management without material stock ownership. Moreover, this management-controlled 44 per cent of the corporations had 58 per cent of the collective wealth of the entire 200. More recent data published by the National Resources Committee⁸ and the Temporary National Economic Committee⁹ appear to indicate that there has been no important shift in the years since 1930.

By means of large investments coupled with full use of minority control devices, a few very powerful interest groups have attained a tremendous degree of power over the corporate structure of the nation, which in turn represents the major portion of economic activity in industry, transportation, utilities, and finance, as well as substantial amounts in other sectors. Grouping the 50 largest financial corporations with the 200 largest nonfinancial, in the above study the National Resources Committee found that out of total assets of \$115,000,000,000 for the group the J. P. Morgan-First National Bank of New York interest group controlled over \$30,000,000,000; and the Kuhn Loeb investment banking interests controlled over \$10,000,000,000, primarily railroads. Of the purely family groups, the Rockefeller interests controlled over \$6,000,000,000, the Mellon interests over \$3,000,000,000, and the du Pont interests over \$2,500,000,000. These figures for assets controlled are of course somewhat tentative, inasmuch as in most cases working control is exercised with much less than a majority of the voting stock. In some cases, in fact, especially in the case of the Kuhn Loeb group, control is exercised primarily without heavy stock holdings and director representation.

⁷ Berle, A. A., and Means, Gardiner C., *The Modern Corporation and Private Property*, Macmillan and Co., New York, 1934, p. 94.

⁸ *The Structure of the American Economy*, Chap. IX.

⁹ *The Distribution of Ownership in the 200 Largest Nonfinancial Corporations*, TNEC Monograph No. 29, 1940.

When it is recognized that the ownership of stock in all American corporations is much less widely spread than is commonly supposed, the concentration of control permitted by the corporate form becomes still more impressive. Only 10,000 persons (0.008 per cent of the population) own one fourth, and 75,000 persons (0.06 per cent of the population) own fully one half of all corporate stock held by individuals in this country.

It should be clear that there is nothing necessarily dishonest or illegal about such centralized control of business enterprises by minority groups. On the contrary, there is strong evidence that in many cases the operation of corporations in this manner leads to far greater efficiency than could possibly be expected were there no such centralization of control. But there are also many cases of the opposite sort, where centralization of control is used by insiders for their own aggrandizement and for the exploitation both of other security holders in the corporation and of those outside the corporation. Centralization of control, especially when coupled with lethargy on the part of other security holders, presents a favorable opportunity for graft and dishonesty by insiders if they desire to operate in that manner. Falsification of reports, payment of very high salaries and fees, manipulation of prices of the corporation's securities, and other such practices have been unfortunately common in the history of American business.

The Cooperative

Cooperatives have been formed both by consumers and by producers in this country. Consumers' cooperatives are organizations in which consumers band together to buy goods from wholesalers or manufacturers and sell directly to themselves, in an attempt to obtain higher quality goods and lower prices than they could through ordinary retail channels. Producers' cooperatives are organizations in which producers band together to sell directly to users of their products, in an attempt to get higher prices and better treatment than they could through ordinary marketing channels. Typically a cooperative is organized as a corporation but it differs from an ordinary corporation in at least two important respects: (1) each cooperative member has only one vote regardless of how many shares of stock he owns; and (2) above a certain fixed dividend per share, any profits are divided on the basis of patronage of cooperative members (that is, purchases from consumers' cooperatives, or sales through producers' cooperatives).

Cooperatives have achieved great importance in England and the Scandinavian countries, but they have developed much more slowly here. At present the country is spotted with consumers' cooperatives, but for the most part these are very small. Producers' cooperatives have developed furthest in agriculture, where unorganized farmers have felt themselves at a great disadvantage in facing organized buyers of their products. At present producers' cooperatives are probably most important

in the marketing of dairy products and citrus fruits. In 1936 there were some 10,500 cooperative marketing associations in agriculture, with a total membership of around 3,000,000 and total volume of business of \$1,840,000,000. Although cooperatives are still restricted to a few fields and remain relatively insignificant, they have grown rapidly in the last two decades, and it may be that in the future they will be of increasing importance.

CHAPTER 5

Size and Concentration in American Enterprise

OF GREAT importance in relation to the way in which economic processes operate are the sizes of business enterprises and the numbers of firms producing each commodity or service. Changes in these things may involve drastic changes in an economy. A brief survey of these characteristics of business enterprise in the United States will stand us in good stead as a background to the analyses which are to follow. It is useful to begin by defining some basic concepts.

Some Basic Economic Concepts

In Chapter 4 we described the most prevalent forms of organization of business units. Business units are frequently called "firms." John Brown, farmer and sole owner and manager of his farm, is the head of a firm that is a single proprietorship. General Motors Corporation with its many ramifications is also a firm. The United States Steel Corporation is a firm with steel mills in many cities, with iron mines, with ore ships on the Great Lakes, and so on. The corner grocery store may be a single proprietorship with just one physical unit, or it may be a member of a large chain organization under one central control with many retail units widely scattered. The important characteristic of the firm is that it is owned and controlled essentially as a unit however diverse its parts.

The function of making fundamental policy decisions in an enterprise is generally called the function of "entrepreneurship." The entrepreneur makes decisions as to how productive agents will be combined, what the quantity of goods produced will be, what the variations in those goods will be, how the concern will be financed or refinanced in times of need, what price policies will be followed, and so on. A firm is thus a business unit under one coordinated "entrepreneurship."

In the independent corner grocery store the proprietor is the "entrepreneur." He decides whether to borrow funds to remodel his store, or what prices to set on his merchandise. It is equally easy to see who performs the functions of the entrepreneur on an independent farm. In more complex forms of business this becomes less obvious. For example, who is the entrepreneur of American Telephone and Telegraph? The 600,000 stockholders? The Board of Directors? The finance committee of the

Board of Directors? The President? It is impossible to pick out any single person or even group of persons as the entrepreneur; the functions of the entrepreneur are performed in a coordinated way by the various individuals and groups concerned.

The firm is the unit within which productive resources of various kinds are combined to turn out goods and services for sale. These resources are frequently referred to in economics as "agents of production" (or "productive agents"). Agents of production are laborers, land, raw materials, machines, buildings, and so forth. They are all the productive resources available in an economy. An "agent of production" is thus a thing or a person that can make a contribution to the production of economic goods and services. Productive agents take literally thousands of forms, and usually in even a single business enterprise a variety of different kinds of agents will be used.

The productive agents employed in a firm include buildings and other more or less fixed physical equipment, which may be located in one or in several different places, each separate grouping constituting a "plant." The term "plant" is most frequently associated with the buildings and equipment of a manufacturing establishment—perhaps a shoe manufacturing plant, or an automobile assembly plant. The Ford Motor Company is clearly a "firm" with "plants" in Dearborn, Seattle, St. Louis, Kansas City, Minneapolis, and so forth. But the Atlantic and Pacific grocery chain is also a "firm" with "plants" in literally thousands of towns across the entire United States. John Brown's farm enterprise, on the other hand, is a "firm" with one "plant" only.

In each physical plant there are employed other agents of production, such as labor and raw materials. The physical plant in combination with these other productive agents is an "establishment." Thus the Dearborn "plant" of the Ford Motor Company is the core of an "establishment" in which hundreds of workers are employed. The local unit of the A. and P. chain is a "plant" which in combination with the shelves of goods for sale and the sales clerks constitutes an "establishment."

It is much more difficult to state what is meant by an "industry" than what is meant by "firm," "plant," or "establishment." For analytical purposes it is frequently convenient to use the term industry to refer to all the producers of a given "commodity." This is, in fact, the definition which we frequently use. Thus Farmer Brown is a part of the "wheat industry" insofar as he produces wheat on his farm. If he produces corn he is a member of the "corn industry." Evidently General Motors is a firm belonging to many different industries. When we try to make our definition of industry very precise we run into many difficulties. Is there a "shoe industry," or a "men's shoe industry," or a "men's dress shoe industry"?—and so on. The definition of an industry set up for any particular purpose will depend on the definition of "a commodity." In some cases "a commodity" may be narrowly defined as one product identical in all respects; on the other hand, for some purposes it will be convenient to broaden the definition somewhat. We therefore defer the precise defini-

tion of an "industry" to each particular situation in which the "industry" is discussed. Here it is possible to define an "industry" broadly as all the producers of a given "commodity."

We may therefore conclude that: the "plant" is a physical unit of production owned by a "firm"; the plant plus the other productive agents employed in the plant is an "establishment"; the "firm" is a business unit under unified ownership and control, which in turn may control one or more establishments; the "industry" is all producers of a given commodity, and will include all "firms" engaged in such production; finally any particular firm may belong to several "industries."

Size of Enterprise in the United States

There are many ways in which we might measure the size of enterprise. When we speak of a firm as "large" we usually mean, among other things, that it receives and expends large sums of money. The "million-dollar corporation" will have invested heavily in plant and equipment for turning out great quantities of goods. In all probability thousands of men are employed and pay rolls are very large. There are usually separate establishments located in various places, all of which taken together make up the total that is the firm. These are the characteristics of big business: large gross money income and outgo, heavy investment in each plant, many employees, and many establishments in combination.

In judging the size of a concern each of these factors and some others might be used as measuring rods. Which measure is chosen in any particular case will depend on the purposes for which the choice is made. If we are most interested in the importance of a firm as an employer of labor, we shall obviously select this aspect of its activities as the measure of its size; if we are interested in the problems created by heavy fixed investments in buildings and equipment we shall be more likely to use evaluations of such investments, and so on.

The size of enterprise differs vastly in different sectors of the economy. Agriculture is by far the most important area of economic activity characterized by small enterprises, both in size of plant and of firm. According to the National Resources Committee, in 1935 "there were nearly 7 million farm units, less than 42,000 of which involved the gainful activity of more than 5 persons."¹ The other most important fields of small enterprise are nonutility services, retail trade, and construction. Most retail enterprises employ only a few workers and operate with small investments in plant and equipment and with small total sales volume. This is still true despite the increasing importance of chain and mail-order organizations in merchandizing. Largest typically are public utility enterprises, and next are enterprises engaged in manufacturing. Yet within any of these groups are important departures from what might be regarded as "typical."

A more concrete picture of sizes of enterprise may be gained by con-

¹ *Structure of the American Economy*, p. 103.

sidering briefly some facts concerning trends in (1) the size of manufacturing establishments, (2) industrial consolidations, (3) the size of enterprises engaged in marketing, (4) the size of agricultural enterprise.

Trends in the size of manufacturing establishments

As a part of its investigation of concentration of power in American industry, the Temporary National Economic Committee provided us with extensive analyses of the scale of manufacturing operations in the United States. Included in this study are data on "size of establishments."

The Committee used several measures of size of establishment; among them were: (1) number of workers, (2) value of output,² and (3) volume of output per establishment. These data are presented in Table 5—1.

1. The number of workers per establishment is found by dividing the total number of workers employed by the number of establishments. The index of workers per establishment is based on these figures of the average number of workers per establishment. The number in each year is figured as a per cent of the number in 1914; the number in 1914 is obviously 100 per cent of itself.

2. The value of output per establishment is simply the sum of the total money incomes of all the establishments each year, divided by the number of establishments.

3. The volume of output per establishment is presented as an index number, since it is impossible to add together bales of cotton, pairs of shoes, bicycles, and so on, and get any meaningful result. The process of computation is too complicated to go into here. The meaning of the index, however, should be reasonably clear. It compares outputs in later years with average outputs in 1914 as a "base year," much as we compared the number of wage earners per establishment with the number in 1914 by the use of percentage figures.

What do these data reveal? Although without detailed consideration and many qualifications not discussed here they are only rough indicators of trends, they do give us a general idea of the number of workers now employed in manufacturing establishments, and of the values of products per establishment. They also tell us something of what happened over the years since 1914. Over this period there was an increase in the average number of workers per establishment, though the short-time fluctuations associated with changing conditions are very large relative to any general increase in number of workers over the whole period. The index of the volume of production per establishment shows a more definite rise over the period, fluctuating in more recent years around a higher level than in the earlier part of the period.

These figures give a slightly exaggerated picture of the average size of enterprise in manufacturing because they omit establishments with gross

² Value of output was expressed both in 1914 dollars and in dollar values of the particular years reported. Because of the complications involved, the "corrected" figures are omitted here.

TABLE 5-1 *

GROWTH IN THE AVERAGE SIZE OF ESTABLISHMENTS IN ALL
MANUFACTURING INDUSTRIES, 1914-1937

(Establishments with products valued at \$5,000 or more)

Year	Average wage earners per establishment (1)		Value of products per establishment (2)	Volume of production per establishment (3)
	Number	Index	Index	Index
1914	37.3	100	100	100
1919	40.1	108	106	104
1921	33.7	90	114	90
1923	42.6	114	154	140
1925	42.8	115	164	153
1927	41.8	112	173	149
1929	40.5	109	177	154
1931	35.9	96	163	123
1933	41.5	111	171	141
1935	42.9	115	172	142
1937	51.4	138	216	185

* Source: TNEC Monograph No. 27, Table 2, p. 4.

revenues under \$5,000. But a much more serious fault is that as averages they fail to show the great variations between different types of manufacture. This deficiency can be remedied by introducing other data. Table 5-2 shows the wide differences in the number of wage earners working in large and in small establishments. Of the 8,569,231 workers employed in manufacturing in 1937, 2,262,282 were in establishments employing over 1,000 workers.

TABLE 5-2 *

NUMBER OF WAGE EARNERS IN ESTABLISHMENTS OF DIFFERENT SIZES, 1937

(Size of establishment measured by number of wage earners)

Wage earners per establishment	Number of wage earners	Percentage of wage earners
1 to 5	170,174	2.0
6 to 20	514,487	6.0
21 to 50	750,922	8.8
51 to 100	852,373	9.9
101 to 250	1,522,670	17.8
251 to 500	1,363,000	15.9
501 to 1,000	1,133,323	13.2
1,001 to 2,500	1,080,534	26.4
2,500 and over	1,181,748	
Total	8,569,231	100.0

* Source: TNEC Monograph No. 27. From tables 3 and 4, pp. 9 and 11.

In which industries do we find the large, in which the small establishments? The Temporary National Economic Committee has listed the eight industries in which establishments averaged over 500 wage earners in 1937. The term "industry" is here used very broadly to include entire groups of commodities. These industries and the number of wage-earners per establishment were: *

TABLE 5—3

INDUSTRIES WITH OVER 500 WAGE EARNERS PER ESTABLISHMENT, 1937

Rubber boots and shoes	1,530
Motor vehicles, not including motorcycles	1,485
Steel-works and rolling-mill products	1,169
Locomotives (railroad, mining, and industrial; not made in railroad repair shops)	692
Copper smelting and refining	631
Cane sugar refining	610
Wool carpets and rugs (other than rag)	560
Linoleum	507

* Source: TNEC Monograph No. 27, p. 14.

There were seven industries in which establishments had an average of fewer than 9 employees in 1937. These were: *

TABLE 5—4

INDUSTRIES WITH FEWER THAN 9 EMPLOYEES PER ESTABLISHMENT, 1937

Cheese	2
Lapidary work	4
Bluing	5
Butter	5
Ice, manufactured	5
Theatrical scenery and stage equipment	8
Vinegar and cider	8

* Source: TNEC Monograph No. 27, p. 15.

Although none of the industries characterized by very small establishments employ a large proportion of the total workers of the country, a large industry employing many workers does not necessarily have very large establishments. For example, the printing and publishing industry and bakeries are both very important in the total economy but employed in 1937 an average of less than 15 workers per establishment.

Industrial consolidations

When combined in one firm are many different establishments all carrying on the same processes, this is commonly called "horizontal integration." An outstanding example is the early combination of oil refineries in the old Standard Oil Company. Modern chain stores are also examples of

"horizontal integration" insofar as they combine a large number of retail units in one "integrated" business. But many business combinations involve integration of establishments at various stages in production processes. The United States Steel Company provides a striking example, with its coal and iron mines, special freight cars, blast furnaces, and final fabrication of many kinds of steel products. Integration under one firm of establishments in these many stages is called "vertical integration." In many cases there is both horizontal and vertical integration, though one or the other aspect of integration may predominate in a particular big concern. Industries characterized by large establishments are frequently characterized also by large firms in which are such "integrations" of a number of establishments. These may be the results of "mergers" of previously separate firms, in which the original concerns lose their identity, and a new concern is created. They may be the result of acquisitions of some concerns by others. Or they may result simply from expansion of a given concern with the creation of new establishments.

A gradual slow development of giant enterprises has taken place over the last half century, but during this time there have been two dramatic spurts of business consolidations, from 1897 to 1903 and from 1925 to 1929. The gradual development seems to be the result of real opportunities for economy plus efforts to gain control of a market. The spurts came in dramatic boom periods, in large part through the actions of financial promoters. The problems associated with these developments will become evident as we proceed with our study, but it is worth noting here that to a considerable extent consolidations have arisen not only because of economic pressure for profits from production and sale of the product, but also because of manipulations of the securities markets by financiers.

The industries which have experienced the greatest consolidation movements are varied, including some kinds of mining, steel manufacture, petroleum refining, tobacco manufacture, sugar refining, and a number of others.

Trends in the size of enterprises in retailing

Retailing is one of the areas of production in which enterprises are still commonly very small. This is obviously true of the size of establishments, though some chain stores are very large concerns when all establishments in the firm are considered together. Table 5—5 gives a classification of retail establishments according to the value of sales in the years 1929, 1933, 1935, and 1939. Surprisingly enough there is actually an increase in the percentage of stores with annual value of sales under \$5,000. The variations between stores are remarkable. As would be expected, the relatively few very large retail establishments do a very large share of total retailing business. This is shown in the second part of the table. In 1935 less than 3 per cent of the retail stores had sales of \$100,000 or over, but they did 38.3 per cent of the total retail business. The average value of sales per store was \$20,000 in 1935, \$32,000 in 1929.

TABLE 5-5*

PROPORTION OF STORES HAVING SPECIFIED ANNUAL SALES AND THE PROPORTION OF TOTAL SALES IN EACH GROUP, UNITED STATES, 1929, 1933, 1935, AND 1939

Annual sales of store	Percentage of							
	Stores				Sales			
	1939	1935	1933	1929	1939	1935	1933	1929
Under \$5,000		41.5		27.2		4.4		2.0
\$ 5,000-\$ 10,000	54.2†	18.5	64.4†	16.5	9.1†	6.5	13.9†	3.7
\$ 10,000-\$ 29,000	29.5	25.6	24.2	31.5	21.3	21.8	24.9	11.6
\$ 30,000-\$ 99,000	12.8	11.7	9.5	19.8	27.3	29.0	28.3	31.5
\$100,000-\$299,000	2.8	2.2	1.6	4.0	18.9	17.6	15.0	20.1
\$300,000 and over	.7	0.5	0.3	1.0	23.4	20.7	17.9	25.1

† All under \$10,000.

* Source: U. S. Census of Business, 1935, *Retail Distribution*, Vol. I, p. 1-31. 1939, *Retail Trade By Sales Size*, p. 2.

Large enterprise in retailing is primarily the large firm composed of many establishments and commonly known as the "chain store." The relative importance of chain stores in the American retail markets is shown in Table 5-6. There are wide variations between different groups of commodities, and in some areas chains are growing in relative importance while in other areas they seem to be declining somewhat. The declines are probably due to a considerable extent to the special taxes which have been imposed on chain stores. Chain stores are most important in the northern and western part of the country, their growth in the South being a more recent development.

TABLE 5-6*

RELATIVE IMPORTANCE OF CHAIN STORES BY SPECIFIED TYPE, 1929, 1933, 1935, AND 1939

Type of store	Percentage of total net sales made by chains			
	1929	1933	1935	1939
Variety	90.1	91.2	90.8	86.8
Grocery without meat	45.7	45.0	38.2	32.4
Shoe	38.0	46.2	50.0	49.7
Filling stations	33.8	35.5	21.5	10.2
Grocery with meat	32.2	43.7	39.1	38.4
Family clothing	27.3	20.3	20.9	17.8
Cigars	25.1	33.9	35.8	27.1
Women's apparel	22.7	23.3	25.2	26.1
Men's clothing and furnishings	21.2	22.0	21.2	22.2
Radio	19.1	15.6	23.1	19.1
Drug (with fountain)	{ 18.5†	{ 25.1†	28.8	27.1
(without fountain)			15.4	20.6
Department	16.7	23.9	26.7	30.0
Furniture	14.2	14.2	13.5	14.6
Restaurant and eating place	13.6	14.9	14.8	13.9
Jewelry	6.4	5.9	9.4	10.5

† All drug stores.

* Source: U. S. Census of Business: *Retail Trade: 1939 "Types of Operation,"* p. 4. 1935, *Retail Distribution*, Vol. IV, p. 9.

Trends in the size of agricultural enterprise

In the stronghold of small enterprise, agriculture, there are some factors that are leading toward larger units. Most important are probably the technological developments of the last three decades. These are felt most notably in the large cash-grain areas of the great plains where agriculture has been revolutionized by the use of the tractor. Yet in spite of these trends, agriculture remains and probably will remain an industry of many relatively small independent firms—individual proprietorships. There are many reasons for this, not the least of which is the management problem involved in controlling many individual geographically scattered farmers under one central organization. Continuing technical limitations on the size of plant and hence on that of the establishment result in limitations on the size of the firm. These and other factors explaining size of enterprise will become more evident as we proceed.

Factors Influencing Size of Enterprise

Why are manufacturing establishments typically larger than those in agriculture? And why are establishments in some sectors of industry typically much larger than in other sectors? Why have big chain organizations in retailing developed so far behind the huge industrial mergers in "heavy industry"? Why do some firms remain small and stick to a limited "line" of goods while others expand and ramify into an elaborate diversity of products? The primary answers to these questions may be classified under four headings: (1) technology, (2) management, (3) market opportunities, and (4) financing opportunities.

Technology and the size of business enterprise

Technological considerations affect the size of business enterprise primarily through their influence in determining the most efficient size of establishment within a firm.

It seems a very obvious fact that technology influences the most efficient size of establishment. In some industries large units are typically more efficient; in others, small units may be economical; but in any industry an establishment can be "too small." There are four principal technological factors which tend to make larger establishments up to a certain point more efficient than small ones, although these advantages do not continue indefinitely: (1) division of labor, (2) economy in the use of expensive equipment, (3) balance of processes, and (4) the utilization of by-products.³

1. The greater the division of labor the greater must be the minimum number of workers necessary for producing a good. Hence closely associated with advantages of occupational specialization are tendencies to

³ This is a simplification of the classification given by E. A. G. Robinson in *The Structure of Competitive Industry*, University Press, Cambridge, England, Revised Edition, 1935, Chapter III.

larger establishments. The more complex a product, the greater the possibilities of saving in this way. From this fact it follows, of course, that the economies of large establishments arising out of division of labor will be greatest in those industries in which the product is complex and the processes of manufacture most easily subdivided.

2. Economy in the use of expensive equipment takes several forms. If a complicated machine operated at capacity can turn out large quantities of a product more efficiently and at lower cost per unit than could many laborers working with less expensive machinery, it will pay to install it. But a small establishment could not afford to use such equipment. It brings a saving only when it is used intensively—it would be quite uneconomical if it were to lie idle a considerable part of the time. Such machines may simply do more efficiently the same thing that was done in one process before; they may perform a part of a process, simplifying the remainder; they may combine many processes in one complex set of operations. Of this last type of machine Mr. Robinson writes: ⁴

The large establishment often differs from the small establishment in having fewer rather than more processes of manufacture. The process of the division of labor is being reversed; one large machine can be designed to take over what has hitherto been done by a series of manual, or less completely mechanical operations. In a small motor factory, panels and wings are often beaten out laboriously by hand. In the large factory an enormous press can be kept busily occupied, supplanting a room full of panel beaters. Another press can turn out in one piece the complete side of the steel motor body, so that a few welders take the place of the many processes of earlier methods of body making. In other cases, two or three or more consecutive processes are performed by a more complicated machine, which thereby eliminates the labor and time required to set up the work on each of the successive earlier machines.

3. What is meant by the “balancing of processes” can be explained by a very simple illustration. Suppose one machine “feeds” another in that the two machines perform the processes of two consecutive stages in the manufacture of a product. In one hour machine A may turn out 100 units ready for machine B. In one hour machine B can take care of 150 units. If we had one of each of these machines, machine B would be kept idle a third of the time. Adding a second machine of the A type would result in excess equipment of A relative to B. But if we use three machines of the A type and two of the B type we can get a continuous flow of 300 units through the two machines. It thus pays to increase the size of the plant to accommodate three A and two B machines. In the complicated dovetailing of the many machines involved in some manufacturing this advantage of balancing processes may lead to the development of a very large plant. A large plant means a large establishment.

4. Closely related to the balancing of processes just described are the economies possible in the utilization of by-products. In a small meat-packing firm, for example, cattle’s hoofs might be so small in quantity as to be regarded as waste. If the size of the firm or establishment were

⁴ *Op. cit.*, p. 25

increased, however, these previously wasted materials might be salvaged. There are many situations in which firms and particular establishments within firms have been expanded in order to make economical use of such by-products.

All of the technological factors mentioned encourage larger sizes of establishment than might otherwise appear. But economies of increasing size do not continue indefinitely. After the division of labor has reached its most efficient point further increases in the numbers of workers employed simply mean duplication of workers. The same thing is true in the use of expensive machines, the balancing of manufacturing processes, and the utilization of by-products. Once the size of establishment (or firm) is great enough to provide these economies no further technological advantages follow from another set of productive agents.

Management and the size of enterprise

Management problems are of primary importance in their effects on the size of an entire firm, as contrasted with technological factors, which primarily affect the efficiency of different sizes of establishments.

The problem of maximizing management efficiency differs in an important respect from that of maximizing technological efficiency. While there may be advantages in greater size through some ranges, eventually a point is reached beyond which not only the management economies of greater size cease, as did technological economies, but even diseconomies set in. The difficulties of management at this point have become so great that returns to effort diminish and costs rise; the larger enterprise may be in this respect at a disadvantage. Our immediate task is to examine those characteristics of management which will lead to economies, and those which will lead to diseconomies, as the size of enterprise is enlarged.

1. First, what are the principal factors which bring increasing economies of management with increasing scale? These fall in two main types: (a) the advantages of specialization and division of labor in management; (b) the advantages of spreading certain types of service of management over a large instead of a small volume of output.

a. Specialization in management is carried to a very refined stage in some large concerns, where purchasing, selling, technical supervision of production processes, financing, are each centered in separate departments with highly trained personnel. This division of functions has the advantages that men with special abilities may be placed where their abilities will be most fully utilized, and that concentration on particular aspects of a complex business will make these men even more expert in their special lines.

b. A number of activities require a certain minimum investment, but may not be much expanded beyond this minimum whatever the size of enterprise. An example of this might be sales forecasting. As the firm is organized to produce a greater output, sales may still be forecast equally efficiently with little increase in expenditure on this service; the

effect will be to decrease the cost, per unit of output, of providing the sales forecasting service. Also a number of activities increase in effectiveness more than proportionally to the increase in expenditure on them. Most obvious among these is research, which may be very lucrative to a large firm which can develop an extensive research staff, but which may not be worth undertaking at all if the cost is to be spread over a small number of units of output.

2. As with technological economies, there is a limit to the economies of management that increasing size may bring. Division of function can be carried only so far as the nature of the business will permit. Beyond that point there will merely be duplication of men performing each service. Moreover, some services of management cannot be efficiently duplicated. It would be very difficult to operate efficiently a concern in which there were two or more parallel heads of purchasing, parallel heads of production, of sales, and so on. Large committees entrusted with the formulation of major policies are likely to be much more cumbersome in operation than are small ones. The larger the firm and the greater the specialization of all activities in the firm, the more essential and the more difficult the coordination of physical processes becomes, the more difficult the reconciliation of conflicting views with regard to major questions of policy. No matter how expertly the large organization is managed, and the various aspects of its activities coordinated, a point will come at which this process becomes too complex—there is too much red tape, there are too many individuals between the manager in the executive office and the worker in the factory or on the road, the structure becomes top heavy and inflexible. Increasing size eventually brings diseconomies of management in large enterprise—costs per unit of output rise.

These diseconomies in the management of large enterprises might be stated in a different way, as the positive advantages of managing a small business. In the small firm major decisions can be made more easily and more quickly; there will be fewer specialists to consult, fewer difficulties in arriving at a mutual understanding of the various considerations pertinent to a problem. This fact gives a special advantage to small firms in fields in which changes must be made frequently, as in the production of style goods. Channels of communication are more direct between worker and management, and supervision is therefore easier. Blind spots, in which wastes are occurring, are less likely to persist. For example, difficulties of supervising widely scattered workers make even so-called large-scale farms small relative to large-scale manufacturing. And along with these organizational advantages there may be some human factors such as greater interest, energy, and enthusiasm on the part of men managing their own businesses, or having a decisive and responsible role to play in a small enterprise.

How these factors leading to both economies and diseconomies of management in large enterprises balance depends on the conditions in the particular industry. Within a given industry there may also be differ-

ences between different concerns, depending on the capabilities and leanings of particular men engaged in managing such enterprises. Individual differences between businessmen will have the effect of exaggerating the economies of management of large enterprise in some instances, those of small, in others.

Market opportunities and the size of business enterprise

The technological and management considerations discussed in the last two sections are of primary significance in their influence on the internal operation of a business enterprise and on the efficiency with which the firm anticipates and adjusts to market conditions. In his search for profits, however, an entrepreneur will be equally interested in obtaining the most advantageous positions possible in the markets in which he operates. These include (1) advantages in the sale of his product, and (2) advantages in the buying or hiring of labor, raw materials, transportation, and other productive services. In each of these respects he would like to be in as strong a bargaining position as possible, and he would like to have an organization adaptable to changing conditions. Each of these points deserves more detailed attention.

1. Small specialty shops or local retail grocers may have certain advantages in selling because of the personal contacts they establish with customers. On the other hand, large firms may spread advertising costs over large outputs so that these costs amount to very little per unit sold. In some kinds of business there is a continuous incentive to expand production of an increasing variety of associated products that may be conveniently sold in the same place. There is also greater security possible in producing a wide range of commodities, instead of putting "all the eggs in one basket." Probably most important of all, there is an incentive to increase the size of the enterprise in order to gain control in the markets in which the products are sold, and through this monopolistic control to obtain unusual profits.

2. The advantages of quantity buying have been discussed perhaps most frequently in connection with large retailing chains. It is generally true that firms in a position to buy in large quantities are at an advantage. It does not cost as much to sell a thousand pounds of soap in one lot as to sell ten pounds each to one hundred different buyers. The lower price possible in quantity buying is thus due in part to the fact that such buying is more efficient. It is also sometimes due to the fact that the large buyer can exert more pressure, sometimes coercive in nature, for especially favored treatment. Such advantages in buying may extend over the entire range of productive services employed in a firm. They may also include advantages in getting lower cost transportation, especially when it involves the differences between shipping in carload lots or in much smaller quantities. Like technological economies, buying advantages do not continue indefinitely with increasing scales of operation.

Vertical integration, by combining a number of stages of production in one enterprise, eliminates many exchanges and thus affects both the obtaining of resources and the disposing of products. Buying is pushed back to first stages and selling is pushed forward toward finished goods markets. Such integration has the advantage that the flow of goods through the various stages in production processes is more readily controlled. Steady supplies of raw materials on the one hand and steady outlets for these raw materials on the other make production planning more certain and less subject to erratic and unpredictable changes. Vertical integration may also facilitate planning in transportation which will avoid some of the wastes of cross-hauling.

Financing and financial manipulation

Although the obtaining of the funds necessary to initiate a very large enterprise may prove difficult, once the large enterprise is established it is likely to have an advantage over the small one in obtaining easy access to further funds whenever needed. This is true both because the costs of making a large loan are less per dollar than for a small loan, and because the prestige and influence of a big business are greater than those of a small business. This advantage is one factor encouraging "bigness." More important, however, is the fact that some people specialize in financial promotion and speculation. These financiers may be in a position to reap substantial gains for themselves by backing major business reorganizations, mergers, and acquisitions. Such ventures may even be detrimental to the efficient operation of the businesses involved; yet men in a position to exploit investors may obtain huge profits though these undertakings.

The Number of Firms in an Industry

One of the most important aspects of size of enterprise is the relation between size and the concentration of an industry in a few firms. Where the production of a commodity is undertaken in hundreds of different firms, the prices of that commodity are arrived at through the free play of individual buying and selling transactions in the markets. These prices are sometimes called "market prices." Examples are wheat and raw cotton prices. Where there are only a few firms, prices are controlled by the administrative decisions of entrepreneurs and are frequently called "administered prices." The wholesale prices of automobiles and typewriters are examples of prices set in this way. Thus when we know something about the number of firms in an industry we have one clue to the nature of the pricing policies that will probably be followed in that industry.

It is convenient to classify production in six major groupings for a brief examination of the degree of concentration of production in a few firms:

(1) government-operated enterprise and the regulated public utilities such as railroads and local power plants, (2) manufacturing, (3) agriculture, (4) mining, (5) construction, and (6) retail distribution and consumer services.

Government-operated enterprise and regulated public utilities

In a very limited number of industries governments in the United States have undertaken production for sale. These are principally the postal services and some public utilities such as water supplies and light and power production. Where government (whether federal or local) has undertaken production it is frequently the only agency providing the service involved; thus control of water supply in a local market is entirely concentrated in the hands of the local government, which sets the prices or rates charged. In many instances, for example railroads, telephone and telegraph, and private local utility companies, government regulates the rates and other policies of the private concern. Occasionally the policies of these firms are determined by the cooperative decision of private groups and government agencies. In any case these services are rather highly concentrated in the hands of one or a few agencies, private or public.

Degrees of concentration in manufacturing

Much greater variation is found in the extent to which manufacturing industries are concentrated in a few firms or spread over many. Data on this subject are included in the *Structure of the American Economy*, a publication to which numerous references have already been made. Table 5—7 gives data concerning the percentage of the value of products in each industry which were sold by the four largest and by the eight largest firms in 1935. These data include only industries employing totals of over 25,000 persons. They have been reproduced in chart form in Figure 5—1.

As presented these data very much understate the degree of concentration in American manufacturing enterprise for two principal reasons:

1. The census in classifying industries necessarily includes under one classification diverse products. For example, "cotton manufacture includes such diverse activities as the making of surgical gauzes, tire fabrics, belting, sheeting, yarns and print cloths."⁵ If the classifications of industries were more refined, the number of firms in each of these narrower classifications would of course be much less. Only when the classification of an industry covers a relatively homogeneous product is the information given by the census data an approximately accurate index to degree of concentration. Otherwise concentration is very much underestimated.

2. These data are for national totals, but many industries produce essentially for local markets. For example, building materials such as cement, concrete, granite are produced by many different firms scattered all over the country; but in any one local market production is usually concentrated in one or a few enterprises, which provide almost the entire local supply.

⁵ *Structure of the American Economy*, p. 115.

What conclusions may we draw from these data? First the great variation in degrees of concentration in different industries is evident. It is also evident that, measuring size of industry by number of persons employed, we find all degrees of concentration in industries of all sizes. But after recognizing this extreme degree of variation, we find some general conclusions of considerable interest. In approximately a third of the census industries producing manufactured goods, four firms are re-

TABLE 5—7

CONCENTRATION IN MANUFACTURING ENTERPRISE, 1935 *

Industry	Value of product: percent of industry	
	Four largest producers	Eight largest producers
A. Industries employing more than 100,000 persons		
Motor vehicles, not including motorcycles	87.0	94.2
Motor vehicle bodies and parts	69.4	76.8
Steel-works and rolling-mill products	49.3	63.8
Electrical machinery, apparatus, and supplies ..	44.4	52.3
Meat-packing, wholesale	55.6	63.5
Railroad repair shops, steam	36.0	53.7
Wool and hair manufactures ..	23.1	32.9
Boots and shoes, other than rubber	24.7	30.8
Canned and dried fruits and vegetables; preserves, jellies, fruit butters, pickles, and sauces	22.7	30.4
Bread and other bakery products	18.0	25.6
Printing and publishing, newspaper and periodical	20.3	25.5
Paper	13.8	21.6
Men's cotton garments	10.6	16.8
Cotton manufactures	8.0	14.0
Machinery, not elsewhere classified ..	6.9	11.0
Men's, youths', and boys' clothing, not elsewhere classified	4.5	7.4
Furniture, including store and office fixtures	5.3	8.8
Knit goods	4.9	8.5
Printing and publishing, book, music, and job ..	4.4	6.5
Lumber and timber products, not elsewhere classified	4.5	7.6
Women's, misses', and children's apparel, not elsewhere classified	1.3	2.2
B. Industries employing 25,000 to 100,000 persons		
Cigarets	89.7	99.4
Rubber tires and inner tubes ..	80.9	90.4
Rayon and allied products ..	74.3	90.2
Tin cans and other tinware, not elsewhere classified	80.1	85.6
Agricultural implements	72.4	87.7
Carpets and rugs	51.1	68.2
Ship and boat building, steel and wooden, including repair work	44.8	64.7
Refrigerators and refrigerating and ice-making apparatus	46.1	58.0

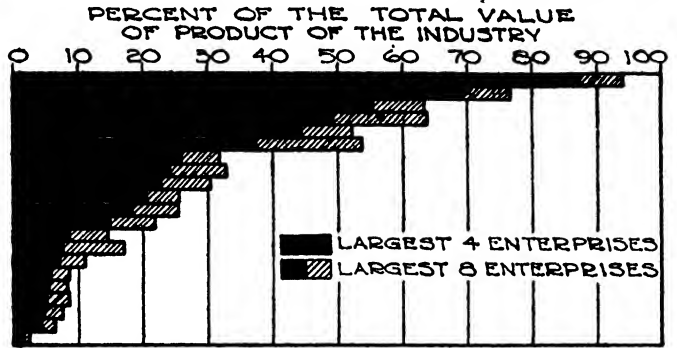
TABLE 5-7 (Continued)

Industry	Value of product: percent of industry	
	Four largest producers	Eight largest producers
Steam and hot-water heating apparatus and steam fittings	38.7	49.2
Petroleum refining	38.2	58.9
Radio apparatus and phonographs	27.0	38.6
Nonferrous-metal alloys, products, except aluminum, not elsewhere classified	36.2	52.4
Glass	44.9	61.0
Chemicals, not elsewhere classified	35.9	48.5
Hardware, not elsewhere classified	36.3	45.5
Cigars	38.5	50.7
Paints, pigments, and varnishes	32.2	41.8
Gas, manufactured, illuminating, and heating	31.2	56.1
Wirework, not elsewhere classified ..	21.7	35.3
Engines, turbines, water wheels, and windmills ...	28.9	47.2
Structural and ornamental metal work, made in plants not operated in connection with rolling mills	24.3	31.5
Pulp (wood and other fiber)	22.7	34.5
Foundries	24.1	32.9
Machine-tool accessories and machinists' precision tools	20.9	30.7
Flour and other grain-mill products	29.1	37.0
Leather, tanned, curried, and finished	21.8	34.3
Rubber goods other than tires, inner tubes, and boots and shoes	18.5	28.5
Rayon manufactures ..	18.6	26.4
Pottery including porcelain ware	18.7	29.1
Drugs and medicines ..	23.4	33.3
Dyeing and finishing cotton, rayon, and silk	13.5	22.2
Paper goods, not elsewhere classified	13.2	23.7
Clay products (other than pottery) and nonclay refractories	19.2	26.6
Machine tools	13.3	23.5
Stoves and ranges (other than electric) and warm-air furnaces	13.1	23.0
Ice, manufactured	19.5	28.0
Boxes, wooden, except cigar boxes	13.2	21.8
Confectionery	12.3	19.9
Silk manufactures	5.5	14.7
Stamped and pressed metal products; enameling japanning, and lacquering	12.0	18.6
Machine shops	7.2	14.6
Liquors, malt	11.8	17.7
Boxes, paper, not elsewhere classified	14.1	20.7
Planing-mill products (including general millwork) made in planing mills not connected with saw-mills ..	4.6	8.1

* Data are from *Structure of the American Economy*, Appendix 7, Table I, pp. 240-242. Industries are listed in an order based on percentage of total persons employed in the industry who are employed by the four largest employers.

sponsible for at least half of the value of the product of the industry; and in sixty per cent of these industries four firms are responsible for at least a third of the total product. If the classification of industries were more

**A. INDUSTRIES
EMPLOYING MORE
THAN 100,000
PERSONS**



**B. INDUSTRIES
EMPLOYING
25,000-100,000
PERSONS**

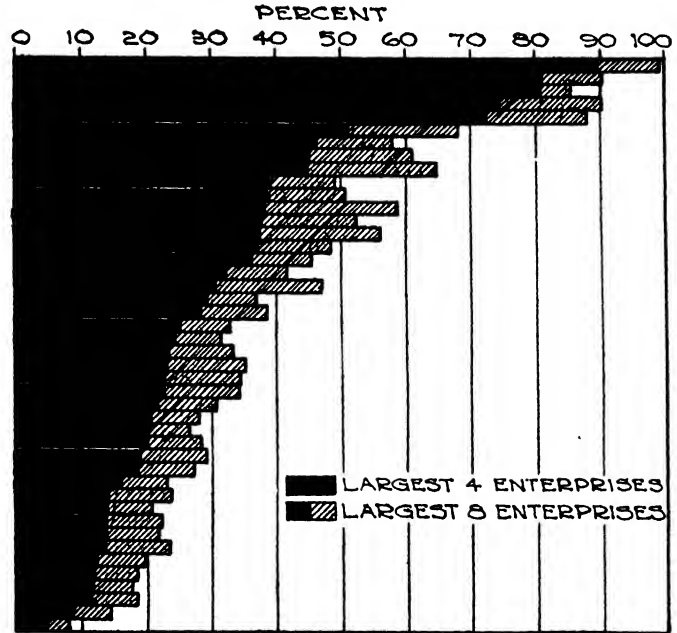


Fig. 5—1. Concentration, measured by value of products, in manufacturing industries, 1935.

refined the degree of concentration in manufacturing enterprise in the United States would appear much greater than this.

Lack of concentration in agriculture

In contrast to the concentration apparent in government and utility services and in some manufacturing enterprises, agriculture is characterized by wide dispersion of control over production. This is especially true of staple crops sold in national markets, products such as corn, wheat,

cotton, hogs, beef cattle. That the largest four producers of cotton, wheat, and hogs produce an insignificant proportion of the total output of their respective industries is shown in Table 5—8. In none of these cases

TABLE 5—8 *

CONCENTRATION IN THE PRODUCTION OF THREE AGRICULTURAL PRODUCTS

Product	Proportion of total produced by		Total number of farms producing
	Four largest producers	Eight largest producers	
Cotton (acres)	0.14%	0.23%	4,850,000
Wheat (bushels)13	.21	1,364,000
Hogs (number)09	.12	3,971,000

* Source: *Structure of the American Economy*, Table IX, p. 116.

do the four largest firms produce more than a fraction of one per cent of the total output of the industry. The contrast between the spread in agriculture and the concentration in industry is dramatically evidenced in this statement: "There were less than 170,000 separate manufacturing enterprises reporting to the census of manufactures in 1935 and making the whole range of manufactured products. Yet there were nearly five million separate producers of cotton."⁶ Cases of a high degree of concentration of control in agricultural production are very exceptional.

Degrees of concentration in mining, construction, and retailing and consumer services

Information concerning degrees of concentration in mining enterprises in the United States is very inadequate. Some mining is known to be very concentrated: iron ore, anthracite coal, bauxite. Soft coal on the other hand is mined on a small scale in many spots over the entire country. No general conclusions concerning the variations in degrees of concentration of mining industries are possible.

There are thousands of small contractors scattered over the United States, and in each urban locality there is usually one or several large contractors. On the whole, contracting is not concentrated, but there is nevertheless a tightly knit structure of controls in the construction industries in many urban markets. In part this reflects concentration in production of some important building materials. In part it is due to the controls exercised by labor groups in the buildings trades. The structure of this group of industries is so complex and in some ways so unique that it is difficult to draw any general conclusions concerning it.

Retailing and consumer services are offered in local markets. When the market areas are small and isolated it is very likely that these services will be concentrated in a few enterprises. In the larger urban centers on the other hand a fairly large number of concerns may be engaged in each

⁶ *Structure of the American Economy*, p. 116.

activity. The development of mail-order businesses has diminished the local monopolies in retailing that previously characterized many small communities.

Relations Between Leadership of the Firm and Diversification of Products

Most firms, whether they are engaged in manufacturing, or agriculture, or retail or wholesale marketing, turn out more than one product. Very large enterprises may diversify into literally hundreds of different items; du Pont, Woolworth, and General Motors are obvious examples. Frequently firms which occupy a major position in one industry are also dominant in others. Some measure of the relation between concentration of control and diversification of production in leading firms is provided by the data presented in Tables 5—9 and 5—10. These data apply to the fifty manufacturing companies having the largest values of products in the year 1937. Together these fifty companies provided 27.7 per cent of the total value of manufactured products in that year.⁷ In addition some of these companies were active in lines other than manufacturing; they engaged also in agriculture, mining, transportation, public utilities, service, finance, and so on. The data in these tables cover their manufacturing activities only.

In Table 5—9 are data concerning the numbers of products manufactured by these fifty largest manufacturing corporations and the proportion of each of these products that was provided by these fifty companies in 1937. Following the classification by industry groups, we find that these companies were much more heavily concentrated in some groups than in others. Their most important relative positions were in transportation equipment, petroleum and coal products, nonferrous metals and their products, rubber products, and iron and steel products not including machinery. Taking all industries together, these fifty companies produced 37.8 per cent of the total value of those products which they manufactured at all. Other data show that in ten of the fifteen industry groups these fifty companies provided at least 20 per cent of total product value. A selection of industry groups according to where the dominant activities of a concern lay gave the following results: 10 of the companies were predominantly active in the iron and steel group, 7 in the food group, 7 in the petroleum and coal group, 5 in the transportation group, and 4 each in the chemicals group, the nonferrous metal group, and the machinery group.

Considerable diversification of production in each concern is evident from the data of Table 5—9. A more specific picture of the numbers of products turned out by each of the fifty big concerns is given in Table 5—10. It must be remembered that a "product" as defined in these reports is still a very heterogeneous thing made up of large numbers of subvarieties. The data therefore understate the degree of diversification

⁷ TNEC Monograph No. 27, p. 583.

TABLE 5--9*

IMPORTANCE OF THE LARGEST 50 MANUFACTURING COMPANIES BY INDUSTRY GROUPS, 1937 (ON A PRODUCT BASIS)

(NOTE: Values represent total production of the products classified in each industry group although some of the products may have been manufactured in establishments classified in other industry groups.)

Industry Groups	Number of companies among largest 50 that manufactured products in each group	Number of products manufactured	Value of products		
			Total United States production	Production by largest 50 companies	Per cent of total
All industries	2,043	44,512,264	16,805,135	37.8
1. Food and kindred products	9	265	8,449,060	2,285,644	27.1
2. Textiles and their products	12	60	1,433,593	71,740	5.0
3. Forest products	23	54	1,324,118	54,784	4.1
4. Paper and allied products	19	38	3,050,967	167,597	5.5
5. Printing, publishing, and allied products†					
6. Chemicals and allied products	38	429	3,110,848	876,441	28.2
7. Products of petroleum and coal	20	43	2,980,408	1,673,996	56.2
8. Rubber products	6	79	895,142	434,639	48.6
9. Leather and its manufactures	5	88	1,147,917	197,197	17.2
10. Stone, clay, and glass products	10	36	558,131	54,102	9.7
11. Iron and steel and their products not including machinery	30	263	7,714,408	3,541,809	45.9
12. Nonferrous metals and their products	19	112	2,025,936	1,135,221	56.0
13. Machinery, not including transportation equipment	25	391	4,430,194	1,292,095	29.2
14. Transportation equipment, air, land and water	15	100	5,772,807	4,114,469	71.3
15. Miscellaneous industries	13	85	1,618,735	905,401	55.9

† Combined with "Paper and allied products group" to avoid disclosing operations of individual companies or establishments.
 * TNEC Monograph No. 27, Table 3, p. 587.

actually existing in the largest manufacturing firms. The data of Table 5—10 are presented in more specific form in Figure 5—2.

TABLE 5—10 *

**DISTRIBUTION OF THE LARGEST 50 COMPANIES BY NUMBER OF PRODUCTS
MANUFACTURED PER COMPANY, 1937**

Products per company	Companies		Products per company	Companies	
	Number	Per cent		Number	Per cent
1 to 5	0	0.0	76 to 80	5	10.0
6 to 10	4	8.0	81 to 85	0	0.0
11 to 15	1	2.0	86 to 90	0	0.0
16 to 20	0	0.0	91 to 95	0	0.0
21 to 25	3	6.0	96 to 100	2	4.0
26 to 30	4	8.0	101 to 125	4	8.0
31 to 35	6	12.0	126 to 150	2	4.0
36 to 40	0	0.0	151 to 175	1	2.0
41 to 45	2	4.0	176 to 200	2	4.0
46 to 50	1	2.0	201 to 225	1	2.0
51 to 55	1	2.0	226 to 250	1	2.0
56 to 60	3	6.0	251 to 275	0	0.0
61 to 65	0	0.0	276 to 300	1	2.0
66 to 70	2	4.0	301 to 325	1	2.0
71 to 75	3	6.0			
				50	100%

* TNEC Monograph No. 27, Table 8, p. 587.

The Causes of Product Diversification

There are many reasons why firms turn out more than one product; many of these reasons are of course closely associated with factors encouraging large enterprise. The discussion of this section will therefore overlap at a number of points that on factors influencing size of enterprise.

Response to diversified customer demands

Sellers frequently seek to carry a "full line" of products that their customers are likely to want. By suiting the convenience of customers in this way, the seller puts himself in a favorable position to make sales and to attract and retain buyers for any one and all of his products. This is true of producers at all stages in the economic system.

1. A manufacturer may diversify in order to provide different items of equipment that tend to be used or purchased together. He may himself initiate additional products as a device to attract customers from competitors. This was the policy of a manufacturer of boxes for the jewelry trade who added jewelers' tags, white display cards, twine, cotton, and other findings.⁸ On the other hand, in some cases a single big customer

⁸ TNEC Monograph No. 27, p. 652.

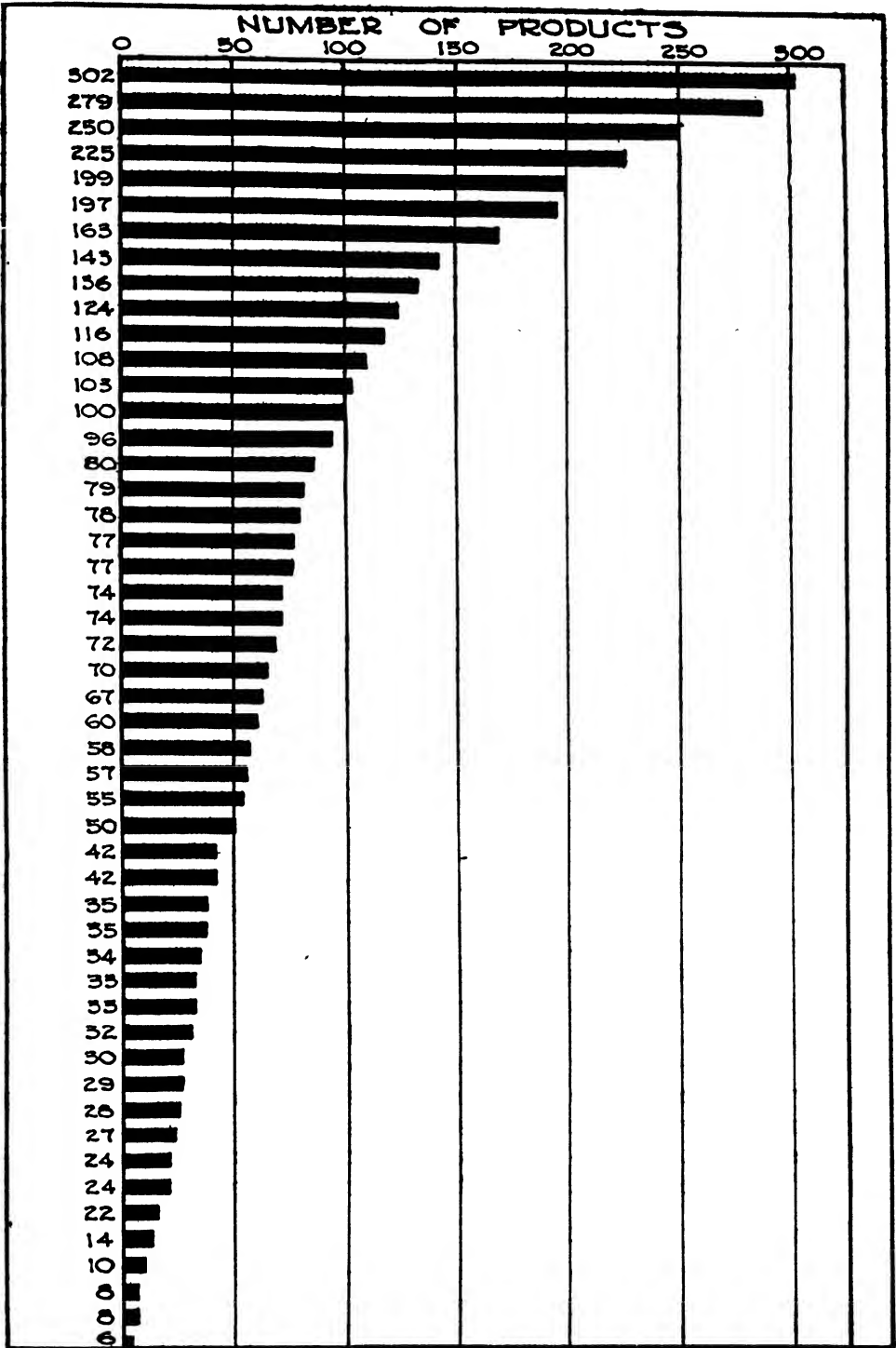


Fig. 5—2. Number of products manufactured by each of the largest 50 manufacturing companies, 1937. *

* TNEC Monograph No. 27, p. 606.

may request special items, and these may be provided in order to retain and encourage his patronage; such special orders may then be made part of a regular line of goods. Large chain stores frequently make such requests, as exemplified by the chain that asked a chemical manufacturer to make dog soap and linoleum paste. The manufacturer responded in order to avoid loss of the chain's patronage in other lines, and the new products were made a part of the regular offerings of the company. Some customer suggestions may be relayed back more indirectly through sales force or through middleman distributors of products. For example, a radio manufacturer added a furniture polish to his line after a number of radio buyers had suggested it, and a manufacturer of auto lubricants and chemicals added tire patch kits in response to requests from customers.⁹

2. A manufacturer sometimes diversifies product because a product which is a main feature of his business will perform more efficiently if auxiliary products used with it are of a certain type. For example, a manufacturer of paper bags went into the production of bag-filling machinery which could be used only with the bag produced by the company. Some important shoe machinery was first introduced by the development of a whole line of auxiliary products including sandpaper, eyelets, nails, tacks, lasts, brushes, dies, and other accessories of a quality that could be used efficiently with the machine. It is not always easy to tell whether diversification justified on these grounds by the manufacturer is in fact a basis of greater monopolistic control over a whole group of products or a straight case of improved performance.

3. Diversification in wholesaling and retailing are so taken for granted that we hardly think of the obvious reasons why diversification appears. Again the retailer responds to the ultimate consumer. Some kinds of goods are frequently bought together (shoes and hose, vegetables and meat), and choice among a variety of goods serving essentially the same purpose is a part of what consumers seek in their shopping. Response to consumer convenience and preferences thus leads retailers to diversify their stocks. This is carried back to the wholesaler who, in response to retailer demands, in turn diversifies his offerings. Indeed it is the primary function of these tradesmen to assemble and offer together in a convenient way the various things their customers wish to buy.

4. A somewhat different case is response to special requests by government. This is of particular importance in a wartime economy.

Efficient utilization of resources

Even more important in the diversification of production is the opportunity provided for more efficient utilization of resources. This opportunity presents itself in many different kinds of situations: (1) efficient use of research facilities, (2) fuller utilization of imperfectly divisible productive agents, (3) efficient use of raw materials in the production of

⁹ *Ibid.*, p. 656.

by-products, (4) efficient use of productive agents in otherwise slack or idle periods.

1. Expenditures on technical research have become increasingly important in manufacturing enterprise in the United States. The advantages are many. The firm first in a field with a new product has a great advantage over others, and may be able to establish a monopoly position through patent controls. The constant changes of a dynamic economy present continuous challenges to each enterprise to be prepared to adapt to change, and research is an important means for realizing such adjustments, especially to find new ways of using old techniques and equipment when and if existing products become unprofitable. While diversification may not be the primary purpose in the building up of such research programs, it is a very likely result. General Motors and Du Pont are extreme examples. Research in these companies has developed both new products based on the use of old techniques and fundamentally new products and techniques. Experimentation with the new products is then followed by production of them, and the resources employed in research are spread over a much larger total of production at a much smaller cost per unit than if these companies were to confine themselves to one or a few products only.

2. The fact that productive agents are imperfectly divisible also provides an incentive to product diversification in some cases. It is this indivisibility that explains some of the economies of large enterprises both in technology and in management. Since an entrepreneur cannot readily split his managers, machines, technicians, traveling salesmen into fractions, he is likely to find some wastes arising in failure to use these agents fully. This may be overcome in part in single-product firms by greatly expanding the size of enterprise so that expensive equipment and personnel may be more fully utilized and so that processes may be more smoothly dovetailed. In many cases, however, it is more convenient and more economical to diversify products in order to make more effective use of these imperfectly divisible resources. For example, a firm manufacturing sound transmission equipment added the production of electric razors and coin-weighing machines because these products all involved precision work and could be made under the same managerial and technical staffs.¹⁰ Variations in the outputs of the different products facilitated full utilization of equipment and personnel. Diversification to make more efficient use of sales clerks in retailing and of the time of salesmen at any level in the economy is typical.

3. That economies may be realized by making use of by-products is one of the facts with which Americans have become quite familiar. This has already been noted as both encouraging and tending to require large-scale enterprise. The re-working of steel filings, the processing of cottonseed oil, the manufacture of fiberboard from the waste screenings in paper mills—these are among the hundreds of examples in American manufac-

¹⁰ *Ibid.*, p. 653.

ture today. Some products initially by-products have gradually taken over major positions in a company. The product diversification involved in a thorough exploitation of these opportunities may, in complex manufacturing processes, divide into literally hundreds of lines.

4. Conditions that may lead personnel and equipment to be idle may give rise to the development of new lines of products that can be produced with these resources. This is sometimes a matter of seasonal slacks, sometimes of decreased demand for or changed techniques in production of the initial major products of a concern. Prefabricators of housing have recently sought to develop a line of products that would enable them to use labor forces and equipment continuously throughout the year. Changing methods of refining oil decreased the need for filtering plants and led a concern originally engaged in filtering operations to diversify by adding the manufacture of industrial lubricants.

Product diversification resulting from adjustments to changes in demand is strikingly illustrated by experience in the textile industry:

A manufacturer of cotton, wool, and rayon textiles believed changes in demand to be a very important cause of the diversification of their product structure. The company does both spinning and weaving and finds it difficult to keep all of both types of equipment busy. A shift in demand that changes the fineness of cloth may be sufficient to upset the balance and make some of the capacity idle. Many commodities have been taken on to utilize the excess spinning or weaving capacity resulting from such changes in demand. A decrease in the demand for an alpaca, for example, caused one company to shift to the production of automobile fabrics. And again, a company began the production of rayon goods when these products began to encroach upon the demand for cotton goods. Some of the added products were substitutes for things previously made of cotton, but some were quite different in nature or use. Changes in styles are often so great that they are responsible for the addition of new products. Practically all the new items of women's wear made by one company were taken on because of changes in styles.¹¹

The quest for security

In some concerns product diversification has been undertaken as a kind of insurance against the risks of economic change, in both production costs and demand. One of the largest chemical concerns in the United States has thus secured for itself a stable position despite erratic demand shifts for some of its products. Such diversification is not necessarily in response to change, but may occur rather in anticipation of the probability of unpredictable change. Diversification for security is of course also a reason for the development of larger enterprises.

Diversification as a corollary of integration

When a concern expands through the acquisition of other concerns, whether in order to develop vertically integrated production or to gain the market advantages of horizontal integration, the newly acquired enterprises frequently produce some products not previously part of the

¹¹ TNEC Monograph No. 27, p. 654.

line of the expanding firm. This line may now be continued and sold to other firms. When a firm integrates its operations directly by the addition of another product to be used by the company itself, it may also sell some of this product outside; such action has the effect of further diversification as well as integration. Examples are many. A manufacturer who needed castings as part of his final product bought out a concern which produced not only castings but other products as well. An enterprise that needed rubber chemicals bought a company producing not only rubber chemicals but non-rubber chemicals also. A company whose principal business was mining, smelting, and selling copper, acquired another company which engaged in jobbing brass and copper products; this step in vertical integration incidentally added screen cloth to the product line since that was produced by the jobbing house. A manufacturer of canned goods started producing cans for his own use and sold some of them to other firms. A producer of heating and plumbing equipment acquired a pig iron manufacturing concern, at first utilizing all this product in the production of the heating and plumbing equipment, but with changed demand and technology it now makes very little use of the pig iron and therefore sells it to others.¹²

Attaining of dominant and controlling market positions

Product diversification may be a tool in building up a tight network of control which will increase the power of the individual concern in dealing with those from whom it buys and to whom it sells, and in eliminating and excluding rivals. The history of meat packing firms provides striking illustrations. In order to establish and maintain controls over the markets in buying livestock and selling processed meats, leading packers spread out into a wide range of grocery products. It is not always easy to distinguish between the serving of customers by diversifying products merely as a response to customer needs, and the use of this technique as a means of gaining control over the market.

Summary and Conclusions

When we add up the varied evidence presented in the preceding pages we get a many-sided picture of business enterprise and concentration in the American economy. Large establishments and firms, and the concentration of activities in relatively few firms in the industry, are most characteristic of government enterprises and public utilities such as railroads and power plants. Manufacturing displays an extremely diverse picture: some industries are characterized by very large enterprises and an extreme concentration of control; others are spread over many firms; and still others fall between these two extremes. Retailing is typically carried on in relatively small establishments located where they will be convenient to consumers; firms engaged in retailing are sometimes quite large, however, because of the combinations of large numbers of retailing

¹² *Ibid.*, pp. 655-656.

outlets in extensive chain organizations. Most generally small in size of enterprise and decentralized and spread in control over production is the general group of agricultural industries.

In examining the reasons why establishments and firms are large or small we focused on four groups of considerations, each of which included several important points: (1) Technological economies arising in (a) division of labor, (b) the use of expensive equipment, (c) balancing of processes, and (d) the utilization of by-products are primarily reasons for at least a fair size of establishment whatever the number of establishments in a firm. These economies continue for larger sizes of establishments in some industries than in others; after a certain point they cease to increase. (2) There are both economies and diseconomies of management as the size of a firm expands. Economies include the advantages which may be gained by specialization of management functions and by the spreading of the costs of some of these management activities over a large number of units of output. Diseconomies arise when red tape becomes overwhelming, when there is a loss of contact between shop and executive office, and when conflicts between those engaged in different management functions become difficult to resolve. The economies and diseconomies of management as the size of a firm expands will depend on the conditions in the particular industry and on the different talents of individuals managing productive enterprises. (3) Especially important in the development of large manufacturing corporations is the opportunity to gain more advantageous buying or selling positions in the markets. Such advantages for the particular firm may mean real increases in efficiency, but frequently they mean the gaining of powerful positions enabling the firm to exercise restrictive controls and exert coercive pressures in the markets in which it operates. (4) Finally, the big waves of expansion in the size of firms through consolidations have come during speculative periods in which financial promoters see opportunities for making profits through promoting such re-organizations. Although large enterprise may mean real savings in financing costs, much of the financial incentive to expansion comes from these groups of financial speculators, and is largely unrelated to efficiency in the carrying on of productive activities.

The differences in size and concentration observed in different industries have important implications for the ways in which resources will be used, the degrees of freedom in moving into and out of particular production areas, and the price, output, quality, and sales policies of the firms involved. The distribution of power in a society, the distribution of income, and the allocation of resources as between different uses, all these and many other important problems are affected by the sizes of enterprise and the degrees of concentration of productive activities.

In most spheres of economic activity there is a tendency toward diversification of the products of a single firm. This is true in agriculture and retailing as well as in manufacturing. By selecting the fifty largest manufacturing companies for an examination of such diversification the

focus was of course placed on situations in which it is most extreme. The causes of product diversification are closely related in many respects to the factors influencing size of enterprise. They include response to diversified customer demands, greater efficiency in the use of resources within an enterprise, the attempt to find greater security for the firm through the development of a number of "lines," the incidental diversification of products associated with integration, and the direct use of product diversification as a means to attain or fortify a position of dominance in the markets in which the firm buys and sells. The importance of product diversification in firms in all areas of economic life is frequently under-rated. In the chapters immediately following it will in fact be ignored, and analysis will proceed on the heroic assumption that each firm produces one good only. Although this assumption is unreal, it permits the viewing of the fundamental operations of the economic system without serious error. As soon as this simplified analysis has been carried far enough to provide an understanding of some basic aspects of economic processes, the fact that there are usually several or many products turned out by a single firm will again be taken into account.

CHAPTER 6

Physical Input-Output Relations in a Firm

THE entrepreneur of a cotton textile manufacturing enterprise and those to whom authority is delegated in such a firm have many kinds of problems which come continually before them. Some of these are technological problems: questions as to how the factory may be organized to get the most efficient production. Some are problems concerning dealings in the markets in which productive agents are bought or hired: questions as to where to buy looms, how to bargain with spinners and weavers and cleaning women, whether to buy raw cotton in large quantities in advance of needs or to wait and purchase it later. Some are problems concerning sales policies: questions concerning variations in quality appeals, in pricing of the yard goods, in pressures to be put on potential customers.

There are many variations in the problems faced by firms in different industries and in different situations; but certain elements in these problems are common to all enterprises—whether they are engaged in cotton manufacture or corn-hog farming or food wholesaling—whether they are dominant in their markets or have close rivals—whether their fortunes are good or bad. These common elements may be first selected for analysis. This procedure involves gross oversimplification, but it provides a background from which to acquire an understanding of the basic factors underlying the organization of productive activity in all its concrete ramifications.

In this chapter attention will be directed briefly to a simplified model of technical possibilities of various combinations of productive agents in a firm. The productive agents used in a firm are called “inputs”; what these agents produce are called “outputs.” Here we shall examine briefly some general characteristics of the relations between “inputs” and “outputs.” Subsequent chapters will explore some simplified models of cost situations and adjustments of firms in their market dealings.

Long- and Short-Run Planning

Producers make plans both as to what to do in the immediate future and as to how to proceed in developing a long-range program. Over short time periods there are only very limited possibilities for increasing or decreasing inputs, for expanding or curtailing outputs. But if time is

given for adjustments, remarkable changes can be wrought. A distinction between "long-run" and "short-run" planning cannot be made entirely clear-cut, but these planning periods may be roughly defined as follows:

The "long run" is a planning period in which the employment of all productive agents is being determined by the entrepreneur. In the long run, new firms may enter an industry and old firms may move out. The long run is no particular time period; it may, in any particular illustration, be a week, a year, a generation, or even a century. The long run in a particular firm or industry is a period long enough so that producers will be able to liquidate even the most durable productive agents such as buildings and machinery, and decide whether or not to re-invest in them; that is, all costs in the long run are "variable" according to the decision of the entrepreneur. The long run for a particular entrepreneur is the period for which planning takes into account complete reconsideration of all investments.

The short run, on the contrary, is a planning period in which some of the productive agents are fixed; there are for this period just so many plants for the production of automobiles, just so many blast furnaces, just so many railroad engines. Entrepreneurs in the industries involved are considering variations in the employment of a part of their productive agents only. They are deciding how many laborers to hire or how much raw material to buy. Again there is no definite time period by which we can identify the short run. And the short run may involve different proportions of variability in agents depending on the particular problem under consideration. If an entrepreneur is considering alterations in production for next week only, he will confine his variations of productive agents used to a very few things, perhaps raw materials and some casual or temporary labor; if he is considering an increase or decrease in the level of operations for the next month, he may vary more agents. The shorter the planning period in any given case, and the smaller the number of productive agents the employment of which will be altered, the greater the proportion which are "fixed." Any situation in which some productive agents are fixed is thus a short-run planning period, though the length of the short run may differ according to the problems under consideration.

The differences between long-run and short-run planning may be clarified by some illustrations. Let us return to the cotton textile manufacturing enterprise once more.

Suppose that a manufacturer of cotton yard goods has already been established in business; he has made investments in buildings and machinery and has been operating his factories for several years. He now sits down at his desk to figure out what should be done over the next month or two. Perhaps this is a period of the year in which he expects to be especially busy: he is just approaching the "peak of the season." He knows that the increased sales of cotton goods that can be anticipated at any given price over the next two or three months are seasonal only, that they will drop back again shortly. How will he adjust to this

situation? Will he build more plant right now in order to take care of this immediate but temporary expansion in production? That is most unlikely. Instead he seeks to hire more spinners and weavers, and he plans to employ his present staff overtime. He plans to keep mechanics on the spot so that any hitch in production through breakdowns of machinery can be quickly corrected. He plans to use more raw cotton. His plans for this two or three months' increase in level of operations involve changes in only a part of the productive agents used in his factories; other agents, such as buildings and machinery, will remain fixed in quantity so far as this production program is concerned. *These are short-run plans.*

Now let us suppose that the cotton textile manufacturing industry seems to be a very profitable one, and that prospects for the future are very bright. Businessmen considering starting in business look at this industry and see in it possibilities of very good returns to their efforts and investments. New firms are then likely to appear in the industry. The men initiating these new enterprises think not in terms of production next month, or even next year, but rather over a long period of years to come. They are planning investments in the construction of plants that will last a long time and be used in production over that extended period. They therefore estimate costs and sales possibilities over a period that extends far into the future. They think in terms of average levels of productive activity in the factory, and they plan both for immediate investments in durable equipment and for continuing outlays in payments to spinners and weavers and other workers and in payments for raw cotton from which to make the cloth. All the productive agents that may be used in manufacturing cotton yard goods are taken into account in their estimates and computations. *These men are engaged in making long-run plans.*

Long-run planning is not, however, confined to the initiators of new enterprises. The already established manufacturer also makes long-run plans at the same time that he must consider short-run adjustments. Although it may last a long time, most productive equipment eventually wears out; and questions arise as to whether or not it would be worth while to replace such equipment. If the industry promises profits for the future, the entrepreneurs of existing enterprises will decide that these agents should be replaced as they wear out, and perhaps that they should be expanded. If the industry seems to be sliding down into a less and less favorable position, if "King Cotton is sick" (as was true in the late 'twenties and the early 'thirties), then durable equipment will not be replaced. When the entrepreneur of an old firm sits down at his desk to figure out what would seem to be advisable concerning replacement of machinery or expansion in the amount of machinery used, when he considers whether buildings should be wrecked, replaced, expanded, he is looking ahead and making estimates of production plans over a fairly prolonged period. If his planning takes into account adjustments in the

quantities of *all* agents used, literally "from the ground up," *then he also is engaged in making long-run plans*:

In drawing a distinction between long- and short-run planning, we have introduced a distinction between "fixed" and "variable" agents of production that will be very useful to us in subsequent discussion. If changes in output for any given time period involve changes in the amount of an agent employed (labor or raw materials for example), then that agent is, *for that period*, a "variable" agent. Whether a particular agent will be "variable" or "fixed" will depend on the planning period under consideration. The cotton manufacturer considering production for next week or for the approaching few months of seasonal peak considers variations in employment of only part of his productive agents, primarily labor and raw materials; these agents are for this period "variable," while his buildings and machinery are for the same period "fixed." If he were considering adjustments in plans for the next two or three years he would consider variations in the machinery purchased and used; in production plans for this period the machinery is a "variable" agent, though buildings may still be regarded as "fixed" since it would not be reasonable to construct or tear down buildings for a mere two-year expansion or contraction of level of operations. If his planning is truly "long-run," he will consider variations in *all* agents used; in a truly "long-run" planning period all agents will be "variable," none will be "fixed" in the plans of the entrepreneur.

A striking example may help point up the social importance of the distinctions between long and short run, between fixed and variable agents, as they are revealed in a critical situation. When a nation shifts from a peace to a wartime basis tremendous shifts in the use of resources are needed. Automobiles must be given up for airplanes and tanks. More steel and magnesium and chromium and copper are needed for defense and must be taken from civilian uses. More shoes and woolen garments are needed for soldiers. But it takes a long time for the war production program really to get under way. The existing aluminum plants may work full-time and overtime, but added workers and materials can carry increased production of aluminum only so far without new plants, which take long periods to be constructed. Short-run increases in the outputs of steel or textiles or airplanes are likewise piddling compared with the expansion that may take place given time for long-range plans to eventuate in new plant facilities and large quantities of products rolling off the new production lines.

Short-Run Combinations of Productive Agents

Suppose Mr. Peterson has just started in business, producing cotton cloth. He has bought some land, put up a building, installed machinery. These agents are now in his employ however fully or incompletely he may use them. If he changes his production plans for the spring months, he will change the number of laborers he will hire for that period, but

he will not change the amount of land and buildings; these have become "fixed" agents so far as planning periods of a few months are concerned. He can vary his output within limits by altering the quantities of variable agents used.

What will happen to Mr. Peterson's output of cotton cloth if he combines various quantities of variable agents with his fixed agents; in other words what will be the relations between "inputs" of variable agents and "outputs" of goods? There are several kinds of questions we might ask. For example: (1) What will be the total output (number of yards of cotton cloth) resulting from employment of any given number of laborers in the factory? (2) How much will one additional man add to the total output at any given point? (3) What will happen to the output per laborer as the number of workers is increased?

If Mr. Peterson were to hire only two or three men they would probably produce scarcely any cotton cloth at all. Each man would have such a variety of tasks to perform that he would be unable to become proficient at any of them, and there would be great wastage of time in his shifting from one task to another. A large proportion of time would probably be required merely to keep the plant in condition; a large proportion of the machines would probably lie idle since two or three men could not watch many machines and keep them going. Materials would frequently be spoiled because no one could have time to concentrate on preventing such mishaps, and because of inadequate inspection. Records would probably be poorly kept and full of mistakes. In other words, the technological and organization requirements of running a factory are such as to require considerable division of labor and specialization within the enterprise if efficiency in operation is to be attained as has already been discussed in previous chapters.

As more men were employed, however, these difficulties would gradually be ironed out, and the output of cotton yard goods would probably rise sharply. Though at first this increase in output would be very marked, the addition of more and more workers would eventually bring smaller and smaller additions to the output of the factory. After there are enough men to make possible an efficient organization of production—to keep machines running, to provide inspection crews, to keep records in order—additional workers would be "worth" less and less to the manufacturer. They would make smaller and smaller additions to the total output of the concern. Though the number of yards of cotton cloth turned out per worker would increase at first, it would eventually decline as greater numbers of workers were employed.

Mr. Peterson would find that employment of "too few" workers would be highly inefficient; he would also decide that whatever his sales opportunities it did not pay to press "too far" in adding workers to operate any given plant because their contributions in additional outputs would be so slight. Just what would be the most profitable level of operation in a given factory would depend both on the physical relations between inputs of labor and outputs of cotton cloth, and on the situations in

the markets in which productive agents (labor, raw materials) are purchased and in the markets in which the goods produced (cotton cloth) are sold.

"Total," "marginal," and "average" product

It would be a very complicated problem to try to investigate thoroughly the possibilities in combining the many different kinds of "variable" productive agents in different proportions with any given "fixed" agents in short-run production programs. In the case of the cotton manufacturer, for example, we should have to take into account different kinds of labor—skilled mechanics, spinners, inspectors, stenographers, cleaning women. We should have to take into account expenditures on office supplies, on fuel, on raw cotton. Moreover, any such detailed investigation would apply strictly to one firm only, and we should have to do the job over again for each firm in each industry. In order to acquire an understanding of basic factors common to all enterprises it is therefore necessary to oversimplify greatly the problem and make some extremely artificial assumptions.

For convenience and simplification in analysis we shall talk about a "unit" of variable agents used as constituting a "unit" of "input." This unit will include a variety of agents such as labor and raw materials, each in certain quantities. Thus one unit of input might be regarded as one "package" of variable agents containing certain set quantities of each of the agents included. The three questions asked about the productivity of labor in the cotton factory may now be posed in a more general form with regard to the effects of adding one "unit" of "input" (one "package" of variable agents): (1) What will be the total output resulting from a given input of variable agents? (2) How much will one additional unit of input add to the total output at any particular point? And (3), what will happen to the output per unit of input as the number of variable agents (units of input) is increased?

The terminology which has been developed to describe physical "input-output" relations in these three aspects is as follows:

1. *Total product* is the total physical output resulting from the application of a given input (quantity of variable agents) to given fixed agents. It will vary with different amounts of variable agents applied, even though fixed agents remain the same.

2. *Marginal product* is the addition (or increment) to total product resulting from increasing by one more unit the input (i.e., quantity of variable agents employed). It is computed by finding the difference between two successive total products.¹

3. *Average product* is output per unit of input. It is computed by dividing total product by the number of units of input of variable agents.

¹ Strictly speaking, it is the addition to output resulting from adding a *very small* unit of input.

Input-output data for a hypothetical firm

The effects of combining different amounts of variable agents with any given fixed agents are most easily studied by the use of a hypothetical example. Let us suppose that Mr. Peterson, having made his investment in land, plant, and basic machine equipment, is experimenting with the results of adding more and more variable agents in order to see what will be the effects on output viewed in its three aspects. The results of his experiment are shown numerically in Table 6—1. An examination of this table reveals the following points:

1. Mr. Peterson finds his maximum total product would be obtained by the application of 21 units of input. The total product increases rapidly at first, especially for the first 5 or 6 units of input. This is to be expected in most enterprises so long as the advantages of specialization and cooperation between productive agents are scarcely tapped. As it approaches more closely its maximum point, however, total product increases more gradually. Finally, when there are so many units of variable agents (in this case 23 or more) that they interfere seriously with efficient operation of the enterprise, total product actually drops. At this point workers are bumping elbows. It is obvious that under no

TABLE 6—1
"INPUT-OUTPUT" DATA FOR A HYPOTHETICAL FIRM

(1) <i>Units of input</i>	(2) <i>Total product</i>	(3) <i>Marginal product</i>	(4) <i>Average product</i>
1	100	100	100
2	350	250	175
3	702	352	234
4	1152	450	288
5	1700	*548	340
6	2190	490	365
7	2604	402	*372
8	2908	304	364
9	3114	206	346
10	3240	126	324
11	3300	60	300
12	3350	50	279
13	3395	45	261
14	3435	40	245
15	3470	35	231
16	3500	30	219
17	3525	25	207
18	3545	20	197
19	3560	15	187
20	3570	10	179
21	*3575	5	170
22	3575	0	163
23	3570	-5	155
24	3560	-10	148

* Maximum figures.

consideration would Mr. Peterson operate his business with so many variable agents as to bring no change in, or to actually decrease, total product.

2. The figures for marginal product (column 3) are found by getting the differences between successive total products from column 2. The maximum marginal product is reached much sooner than in the case of maximum total product, at 5 instead of 21 units of input. From 5 units of input on, marginal product is declining. The fact that the twenty-second unit makes no changes in total product is shown by a zero marginal product; the actual decrease in total product beyond that point is shown by a negative marginal product. It is again evident that less than 22 units of input will be employed in Mr. Peterson's enterprise.

3. The figures for average product (column 4) are found by dividing total product (column 2) by units of input (column 1). The change in average product is, as would be expected, much more gradual than in marginal product, and it never falls below zero. The maximum average product is reached with the employment of 7 units of input. From this point on average product is declining.

Graphic presentation of input-output data

The interrelations between the output data in Table 6—1 may be illuminated by the use of graphs.

Total product is presented in Figure 6—1. Inputs are marked off along the base line; the associated total products are represented by the height of the bars. The first rapid and then more slowly increasing total products are readily seen, as is the downward turn which would appear eventually if inputs were carried far enough. The entrepreneur would obviously operate somewhere short of 22 units of input.

Marginal product can also be seen in Figure 6—1. It is the vertical distance between successive bars representing total product. Where this distance is greatest (at 5 units of input) marginal product is at a maximum. Where total product actually drops off, the vertical distance between successive bars indicates the degree to which additions to inputs would actually cut down output, and thus measures the size of negative marginal products. Between 21 and 22 units of input there is no change in total product, the height of the bars is the same, and marginal product is zero.

Marginal product is plotted directly in Figure 6—2, being represented by the height of the solid bars. It is easy to make comparisons between this graph and Figure 6—1. The highest bar in Figure 6—2 is at 5 units of input, corresponding with the input at which the vertical distance between successive bars in Figure 6—1 was greatest. Total product, represented in Figure 6—1, stays at the same level (at a peak) between 21 and 22 units of input, when marginal product becomes zero. And negative marginal products which appear at 23 units of input and beyond, are represented in Figure 6—2 by bars drawn below the base line, paralleling the drop in height of bars for total product in Figure 6—1.

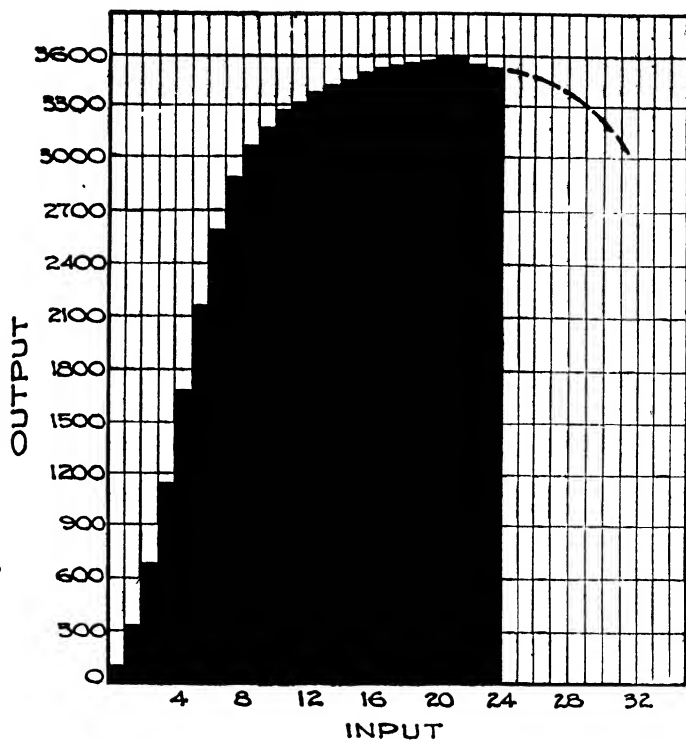


Fig. 6—1. Total product.

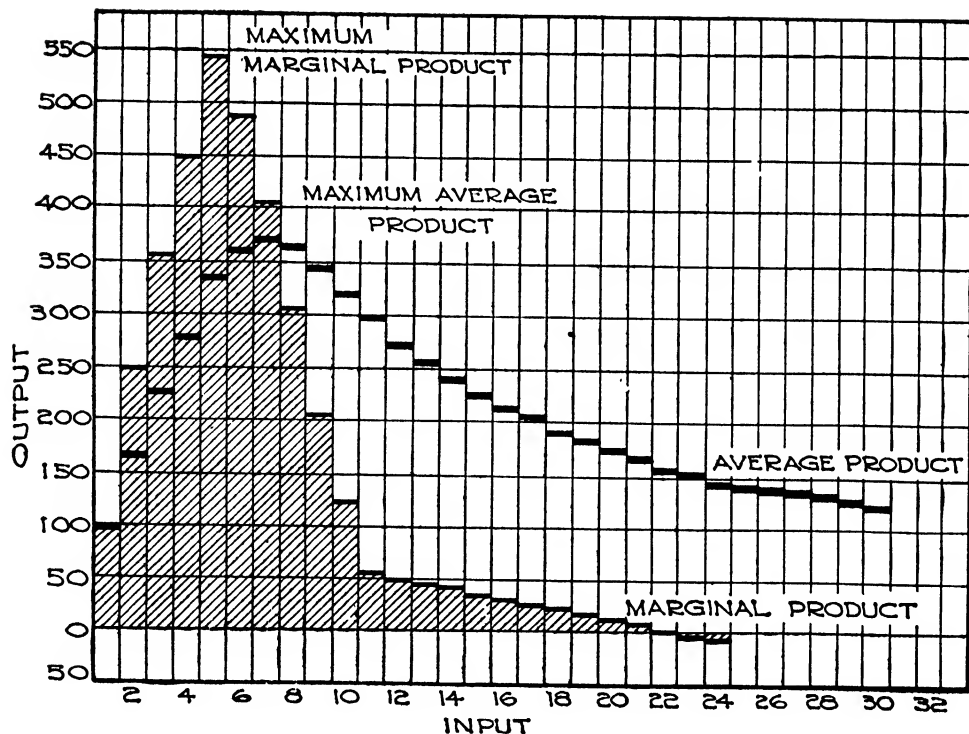


Fig. 6—2. Marginal and average product.

The heavy horizontal lines in Figure 6—2 represent average product. The greater evenness of average as contrasted with marginal product is immediately evident. Another significant relation also becomes clear. So long as marginal product is higher than average product, average product rises; when marginal product becomes less than average product, average product starts diminishing. This must always be the case, and for a very simple mathematical reason. It will be remembered that marginal product is in fact the amounts added to total product by each addition of a unit of input, and that average product is total product divided by the number of units of input. If we add to a sum of numbers (total product) a figure (marginal product) greater than the average (average product), it will pull the average up. For example, the average product at 5 units of input is 340 (which was $1,700 \div 5$); the addition of the sixth unit of input increases total product by 490 (from 1,700 to 2,190), a marginal product greater than the average product from 5 units of input; the average product resulting from 6 units of input must therefore be greater than the average from 5 units of input, though still less than the marginal product from the sixth unit; average product at 6 units of input is in fact 365 (which is $2,190 \div 6$). In other words, whenever marginal product is higher than average product, the added unit of input will pull the average product up, and this will occur even though each successive amount added is growing smaller (i.e., even though marginal product is declining). Conversely, as soon as marginal product is less than average product, the average will be pulled down.

Long-Run Combinations of Productive Agents

In long-run planning an entrepreneur considers the technical possibilities not only of varying the quantities of some agents in combination with a fixed set of other agents, but possible variations in all productive agents. Mr. Peterson may decide to expand his plant, buy more land, and buy or rent more machines; or he may decide that his enterprise is too large and that it would be better not to replace some of his machines as they wear out. These decisions will depend on many things, but one of the facts he will take into account is the different outputs he could obtain when combining relatively variable agents, such as labor and raw materials, with different sets of relatively fixed agents such as buildings.

It will usually appear that the greater the investment in relatively durable agents such as buildings, the more labor and raw materials can be efficiently used. There would be a different set of short-run input-output data for each different investment in the relatively fixed agents. But however he looks at the problem certain generalizations arise out of technological factors. If with any given quantity of one agent are combined larger and larger quantities of other agents, outputs will show the characteristics described in the preceding section. This applies if we consider a given number of laborers and the combining of larger and

larger investments in buildings and machines with these laborers, just as in experimenting with combining larger and larger numbers of laborers with a given physical plant. *Long-run planning will take into account these varying technical possibilities of all kinds of variations in the combinations of all productive agents.*

CHAPTER 7

Costs of Production

AN ENTREPRENEUR combining productive agents in various ways is seeking, if possible, to obtain revenues greater than costs. A brief examination of the nature of costs is therefore essential to an understanding of the functioning of a firm.

The Meaning of "Costs"

"Cost" may mean many different things but for purposes of analytical discussion it is convenient to adopt standardized terminology and to employ concepts which will be of the greatest analytical usefulness. We shall, therefore, define costs as those payments necessary to attract any given number of productive agents and quantity of investment funds into a firm. The chief problem which arises in defining costs is the problem of distinguishing between costs and profits. In order to clarify the economic definition of costs which we have just roughly stated, we shall undertake: (1) a brief examination of what the component parts of costs as we have defined them are; (2) a comparison of a manufacturing corporation's view of costs with ours; and (3) a comparison of an independent farmer's view of costs with ours.

The "explicit" and "implicit" components of costs

In some cases an entrepreneur enters into a contract with an owner of a productive agent to pay him a certain amount. He contracts, formally or informally, to pay wages to a laborer for his labor; he contracts to pay rent to the owner of land for the use of his land, or a royalty to the owner of a machine for the use of that machine. He contracts to pay a cotton trader for the raw cotton he buys, and a machine producer for machines he purchases. He also contracts to pay interest to the individual or the bank loaning him money. Some of these promises to pay are actually written out; others are verbal agreements. All of these payments are costs to the firm entering into the agreements, and the amounts of the payments to be made are in each case made "explicit," whether in written or oral form. They may therefore be described as *explicit costs*.

Some of the funds and the productive agents used in a business, however, are provided by the entrepreneurs or owners of the business themselves. A businessman owning and heading up a small manufacturing

enterprise does not generally make a formal agreement with himself that he will pay himself a certain amount in return for his own labor or as rent for his own land or as interest on his own savings invested in the enterprise. But the labor and land and investment funds he provides nevertheless involve a kind of "cost" of production in his enterprise. He will not ordinarily use his labor and land and funds in this way unless he anticipates that he will get back in return at least as much as he could get in any other way. Similarly, investors in corporate stocks are owners of the corporation; they do not make any explicit agreements with themselves as to what they will be paid in interest on the funds they invest. Yet they must expect to get at least as much as appears to be available to them in other investments or they will not place their funds in this corporation. There is *implied* in the actions of proprietors who put their own labor into an enterprise, anticipation of a return for that labor at least as great as they could get by offering their labor services elsewhere or engaging in other kinds of business undertakings. There is *implied* in the actions of those who invest as owners in a business an anticipation of return on their funds at least as great as could be obtained in alternative investments. The lowest amount that must be anticipated to be sufficient to attract the labor of entrepreneurs, or the land or the investment funds of owner-investors is just as truly a cost of conducting a business as are the wages agreed upon in the hiring of labor, the rents for the hiring of land, or the interest on borrowed funds. These costs are "implied wages," "implied rents," "implied interest." They may be conveniently designated as *implicit costs*. Popular recognition of the nature of implicit costs is expressed in the oft-heard comment "I'm losing money at this," when the speaker means that he is not getting back as much for his labor, or land, or investment funds here as he could get in some kind of alternative employment, and that he would not go into this business again if the choice were open to him.

It is important to remember that both implicit and explicit costs are anticipated only; the amounts anticipated may or may not be recouped. Yet these anticipated payments taken all together are the total costs in an enterprise. Any excess income received by the firm beyond that necessary to cover these costs is "profit" in the economic sense of the term. Any deficit is a "loss," frequently called a "negative profit."

"Costs" and "profits" as viewed by a manufacturing corporation

Corporate accounting uses the words "cost" and "profit" in ways which differ from the meanings which we have just presented. It includes as costs only explicit costs.

A corporation buys materials for use in production; the sums it pays for these materials become the gross incomes of those firms selling the materials, but to the corporation doing the buying these payments clearly constitute a cost. The corporation also employs laborers, and the pay rolls which constitute the income of the laborers are a cost of operation of the business. These "costs" would be among the sums which are listed

as operating expenses in the corporation's accounts. Besides paying for the services of agents such as labor and raw materials, the corporation must pay for its plant and equipment. If an entrepreneur has invested in building and equipment that he expects to use over a period of, say, ten years, he does not expect to cover the entire cost of this equipment by one year's sales of goods. He will allocate the cost of this investment over the ten years of its life, and the share allocated to any particular year will be recorded as a cost of "depreciation" for that year. Another important item which will appear as a cost in the corporate statement is interest on bonds. This may be regarded as a payment for the service of providing funds to finance the enterprise; it is universally considered a "cost."

Thus far we have encountered no real difficulties. The costs described appear in the corporate profit and loss statement as expenses of the enterprise, and are entered for a given time period. They are all explicit costs. The remainder of the gross income of the corporation is either paid out to stockholders as dividends or is reinvested, thus increasing the values of their holdings. Does this income which is held collectively by the stockholders or is paid out to them constitute "costs" or "profits"? It is reported as "profits" by the corporation, but this practice introduces some curious analytical inconsistencies when viewed in more general terms.

Suppose there are two corporations, which we shall designate as A and B, each of which has a total sum of \$1,000,000 invested partly in stocks and partly in bonds. Of the \$1,000,000 investment in Company A, \$700,000 is 5% bonds and \$300,000 is stocks; in Company B \$700,000 is stocks and \$300,000 is 5% bonds. If the return in both corporations, after covering other costs, was \$50,000, or exactly 5 per cent on the \$1,000,000 investment how would this be recorded on the income statement? In Company A 5 per cent of \$700,000, or \$35,000 would be called "costs," \$15,000 "profits." In Company B 5 per cent of \$300,000, or \$15,000 would be called "costs," and \$35,000 "profits." Yet Company B earned exactly the same returns on total investment as did Company A. In general, a corporation with a large part of its financing in stocks and very little in bonds might have exactly the same expenses otherwise and exactly the same gross income as a corporation with a large percentage of bonds; yet the corporation with a large percentage of stock would appear to have a greater total profit than that with more bonds. Thus an arbitrary distinction between costs and corporate profits appears, and the amount of profits so defined depends on the particular way in which a corporation carried out its long-term financing.

In economic analysis, only that part of the income to stockholders which is in excess of a "normal" rate of return would be regarded as "profits." The reason for this is not difficult to see. The stockholder like the bondholder performs a service in providing funds for the financing of this enterprise. To the extent that the payment he receives is necessary to persuade him to make this contribution it is a cost, in much the

same way as is the interest payment to the bondholder. Unless he anticipates receiving dividends here, he will invest his funds elsewhere. Up to the rate of interest generally available to investors, dividends on stock are thus "implicit interest." The investor in stock will look for at least the "going rate of interest" that he can obtain in alternative opportunities. This rate of interest is a cost; it is that amount the anticipation of which in dividends is necessary to attract the investment in stocks—quite as the interest on bonds was necessary to attract bondholders. If there is any income over and above that necessary to attract and retain investment in these stocks, the excess is a "profit" in the economic sense. If returns are insufficient to cover the "implicit interest" to the stockholders, the discrepancy is a "loss."

Let us return once more to the case of companies A and B illustrated above. Assuming the "going rate of interest" to be 5 per cent, which is also the rate of interest stated on the face of the bonds, what may we say of "economic profits" in these two concerns? We shall take three situations: (1) We shall suppose both firms to earn \$60,000 after covering all costs except interest payments, (2) We shall suppose both firms to earn \$50,000 after covering all costs except interest payments. (3) We shall assume both firms to earn \$40,000 after covering all costs except interest payments.

1. On the assumption that both corporations make \$60,000 as a return on the million dollar investment, each will realize \$10,000 in economic profit; \$50,000 of the \$60,000 is a cost. Company A will pay out to bondholders 5 per cent interest on the \$700,000 they have invested, or the \$35,000 indicated above. The remaining \$25,000 it will report as "profits" to stockholders; but \$15,000 of this is the implicit interest (at 5 per cent), anticipation of which was necessary to attract the \$300,000 of investments in stock. The \$25,000 accruing to stockholders in Company A is thus partly economic cost, partly economic profit; to be exact it is \$15,000 implicit interest and \$10,000 economic profit. Company B will pay out to bondholders 5 per cent interest on an investment of \$300,000, or \$15,000. This leaves \$45,000 accruing to stockholders, and the entire \$45,000 will be called "profits" by the corporation. But \$35,000 of this just covers implicit interest at a "going rate of interest" of 5 per cent on the \$700,000 invested in stock; the other \$10,000 is economic profit.

2. If the returns in these two corporations had been \$50,000, at a going rate of interest of 5 per cent this return would just cover the explicit costs of 5 per cent interest to bondholders in each concern plus 5 per cent implicit interest on investments in stock. Economic costs are then just covered in each firm; neither is making any economic profits, and neither is suffering economic losses.

3. If returns in these corporations were only \$40,000 there would be an economic loss (or "negative profit") of \$10,000 in both cases. After pay-

ing \$35,000 to its bondholders, Company A would have left only \$5,000 for stockholders; this is short of a 5 per cent return on the \$300,000 investment in stock. The implicit interest of \$15,000 on stock in Company A is not covered; the deficit of \$10,000 is an economic loss even though the corporation would report \$5,000 profits. After paying \$15,000 to its bondholders, Company B would have left only \$25,000 for stockholders; this is again short of a 5 per cent return on investments in stock. The implicit interest of \$35,000 on stock in Company B is not covered; the deficit of \$10,000 is an economic loss even though the corporation would report \$25,000 profits.

Costs and profits as viewed by an independent farmer

The farmer, like the entrepreneur of a large manufacturing corporation, receives a flow of income from the sale of his product, which, for any given time period, is called the "gross income" of his farming enterprise for that period. Like the entrepreneur of a corporation, he also makes payments to productive agents that he hires or buys, to farm laborers, to farm implement manufacturers, and so on. He pays interest to the bank on the loans which have been extended to him. Usually he provides sizeable sums himself that he invests in addition to the funds that he borrows, just as the stockholder-owners of a corporation provide funds for the furtherance of their business. If the farmer were to invest his savings elsewhere he would receive interest. Some of his earnings must therefore be regarded as an "implicit interest" just as were the normal dividends of the stockholder.

The case of the farmer introduces a new problem in the definition of costs and profits. He is entrepreneur-owner-worker. The corporation hired its executives, and their salaries were reported as a part of the expenses of the business. The farmer receives whatever remains from his gross income after paying all creditors. This remainder is in part a payment for his work as manager-laborer; it is in part a payment of interest on his own investment or rent on his own land. If the farmer is assumed to be acting to maximize his own income in all its parts, he will not put his time, effort, and money into farming if alternative opportunities offer greater rewards. The "implicit interest" on his investment is the cost of attracting his own investment to this enterprise. His "implicit entrepreneurial wage" is the cost of attracting his own labor into this venture. If there is anything remaining beyond the sums properly allocated to these purposes, it is an excess residual, i.e., a true economic "profit."

Recognition of the economic definition of costs is showing up increasingly in cost-accounting practices of farmers, and this is the basis on which the account books kept under supervision of the extension service are set up. Only by making a careful evaluation of what he might otherwise get for his own labor and investment does the farmer arrive at an accurate picture of how well or how poorly he is getting along in his present activities.

A Short-Run View of Costs

Long-run adjustments involve planning with regard to all productive agents; all costs are therefore within the purview of the entrepreneur in long-run planning. In the short run, however, some costs have already been incurred and will go on for a while whatever the entrepreneur of an enterprise may do, while others can be varied in the making of short-run plans. In the short run the distinction between "fixed" and "variable" costs is important and has far-reaching implications with regard to the policies of the firm. The distinction is similar to the classification of productive agents as "fixed" and "variable." We shall examine this distinction in costs and its implications briefly.

The short-run distinction between fixed and variable costs

Fixed costs are costs that do not change with changes in output over a particular planning period. They may be expressed in per annum or per month figures. The depreciation on equipment insofar as it occurs independently of whether or not the equipment is used is thus a fixed cost for a going concern until time comes for replacement. So also is the interest on the investment of bondholders in the corporation, or of the bank in the farm. Fixed costs may also include the "implicit interest" to the owner investor, and the "implicit entrepreneurial wages" of a proprietor of a factory or store or farm. It is important to remember that costs of any sort are not necessarily met, that "fixed costs" are fixed only in that they do not change with changes in output in a particular planning period. In fact, "fixed" costs generally include implicit interest and they frequently include implicit entrepreneurial wages. If the gross revenues of a firm are insufficient to cover all costs these fixed implicit costs are the first type of cost payments that will not be met. And fixed costs in explicit interest on long-term investments by lenders, say bondholders, will usually be the next group of costs to be unpaid when fortunes of the firm are bad.

Variable costs are costs that will be varied with changes in output over a particular planning period; they are the sums paid for the use of variable agents (and as interest on short-term loans) for a given period of time. These costs will be greater or smaller depending on the number of variable agents employed. Since changes in output can be brought about within a given time period only by increasing or decreasing the number of variable agents employed, the aggregate cost of these agents must vary when output varies.

The significance of this distinction between fixed and variable costs is easily illustrated. To return to Mr. Peterson, the cotton cloth manufacturer visualized in the previous chapter, let us suppose that he has already made investments in buildings and machinery. These investments involved the incurring of costs in the past. The costs were partly in payments to contractors for construction of factory buildings, partly in payments to machine producers for the looms and spindles purchased. They also include commitments to pay explicit interest to those making long-term loans to finance these activities, and anticipations in implicit

interest on owner investments. These are "fixed" costs so far as day-to-day or month-to-month adjustments in this cotton factory are concerned. By estimating the life of the machines and buildings the costs of these productive agents may be allocated on a per month basis; let us suppose that the per month fixed cost of these investments is \$800 in depreciation allowance and \$200 in monthly interest costs. Mr. Peterson may decide to step up production next month; but in making such an adjustment he will not put any more money into buildings and machines, he will not incur any further commitments for interest payments on long-term investments. These costs are "fixed" so far as short-run changes in output plans are concerned; they will be the same whether he turns out 100,000 or 50,000 yards of cotton cloth per week during the next month.

Mr. Peterson's decision to increase or decrease the level of output over the next month will affect some of his costs, however. If he decides to increase outputs he will have to use more raw cotton, thus increasing his raw material costs; he will hire more workers for longer hours, thus increasing his pay rolls; he will incur costs of interest on short-term loans negotiated in order to meet these higher pay roll and raw material bills. His short-term decisions will therefore require a close examination of the relation between these *variable* costs and the outputs and revenues which he may obtain. If he decides to increase the level of output of cotton cloth these variable costs will be greater; if he decides to curtail operations somewhat, these variable costs will be cut.

In Table 7—1 are some cost data for a hypothetical firm with input-

TABLE 7—1

COST-OUTPUT DATA FOR A HYPOTHETICAL FIRM

Fixed cost: \$1000
Costs per unit of input: \$100

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Units of input</i>	<i>Total (output)</i>	<i>Average product</i>	<i>Fixed costs</i>	<i>Variable costs</i>	<i>Total costs</i>	<i>Fixed unit cost</i>	<i>Variable unit cost</i>	<i>Total unit cost</i>
1	100	100	\$1000	\$ 100	\$1100	\$10.00	\$1.00	\$11.00
2	350	175	1000	200	1200	2.86	.57	3.43
3	702	234	1000	300	1300	1.42	.43	1.85
4	1152	288	1000	400	1400	.87	.35	1.22
5	1700	340	1000	500	1500	.59	.29	.88
6	2190	365	1000	600	1600	.46	.274	.73
7	2604	372*	1000	700	1700	.38	.269**	.65
8	2908	364	1000	800	1800	.34	.275	.62
9	3114	346	1000	900	1900	.32	.29	.61**
10	3240	324	1000	1000	2000	.31	.31	.62
11	3300	300	1000	1100	2100	.30	.33	.63
12	3350	279	1000	1200	2200	.30	.36	.66
13	3395	261	1000	1300	2300	.29	.38	.67
14	3435	245	1000	1400	2400	.29	.41	.70
15	3470	231	1000	1500	2500	.29	.43	.72

* Points emphasized in discussing physical outputs from varying agent combinations.

** Critical points in cost data.

output figures like those of Table 6—1; the first three columns of input and output figures are in fact copied from that table. Since this is a short-run situation some costs are fixed, some variable; and fixed costs remain unchanged whatever the input and output of the firm may be. Variable costs (column 5) are larger the larger the input-output level contemplated. Since total costs are a sum of fixed and variable costs, they rise with the rise in variable costs. This is essentially the kind of situation that would be faced by an entrepreneur contemplating short-term adjustments in operations; he might be a cotton cloth manufacturer, or a farmer, or an entrepreneur in any other type of business.

Total, fixed, and variable costs per unit of output

Columns 7, 8, and 9 of Table 7—1 have not yet been explained. Total unit cost is total cost per unit of output; variable unit cost is variable cost per unit of output; and fixed unit cost is fixed cost per unit of output. Thus fixed unit costs (column 7) are found by dividing fixed costs (column 4) by total product (column 2); variable unit costs (column 8) are found by dividing variable costs (column 5) by total product; total unit costs (column 9) are found by dividing total costs (column 6) by total product.

The headings of columns 7, 8, and 9 are not as unfamiliar as they may seem. When we ask how much it costs to produce a Ford car we are really asking what the "average cost," or "total unit cost," is at some reasonable level of operation of the enterprise; we are not asking what the total expenses of running the Ford Motor Company are. Similarly, when we speak of the "labor cost" of producing a Ford we are interested in that portion of costs per automobile that may be attributed to labor; labor is one element in variable cost, and in quoting it per automobile we are approximating the concept of "variable unit cost." Finally, when we talk of the part of the cost of producing a Ford that is an "overhead cost" we are referring to fixed unit costs in car manufacturing.

There are several interesting and important characteristics of these cost data that require careful attention:

1. *Fixed unit costs grow steadily smaller the greater the output.* This is necessarily true since we are dividing the same fixed cost figure (of \$1,000) by larger and larger outputs. Clearly the Ford Motor Company would find its production costs per Ford extremely high if its vast plant facilities with all their efficiency were used to produce only a few hundred cars a year. The tremendous cost tied up in plant would be divided among so few cars that the cost per car, the "fixed unit cost," would be extremely high. Fixed costs per Ford are low only when many cars stream off the assembly lines. Where fixed costs are not so large, the importance of the large output in reducing fixed costs per unit produced is of course not as great, but the general relation between output and fixed unit cost is nevertheless a universal one for all types of business.

2. *Variable unit costs first drop and then rise again.* The reason for

this is a little less obvious than is the explanation of the continuous decline in fixed unit cost with expanding production. The greater the output *per* \$100 in costs the less will be the cost per unit of output. If a "unit of input" ("package of variable agents") costs \$100 whatever the number of units of input (the assumption made in Table 7—1) then the greater the product per unit of input the less will be the variable cost per unit of output of the good or service produced. Output per unit of input is what we have termed "average product." When average product increases we would then expect variable unit cost to decrease. The labor and raw material and other variable costs *per* pair of shoes will be less the greater the number of pairs of shoes turned out per laborer and per calf skin. A comparison between columns 3 and 5 of Table 7—1 shows this relationship clearly. Since, as was shown in Chapter 6, *average product rises and then falls, variable unit cost must fall and then rise.*

3. *Total unit costs are equal to the sums of fixed unit costs and variable unit costs at each input-output level.* This is essentially saying that the overhead cost per Ford car plus the costs of operating the plant per car equals the total manufacturing cost per car, or that the real estate and machine cost per pair of shoes plus the labor and raw material cost per pair of shoes equals the total cost per pair of shoes. We have merely classified costs in two categories, "fixed" and "variable." The sum of these two parts equals the whole; and the sum of these parts expressed per unit of output equals the total per unit of output.

4. *Total unit costs fall at first and start rising again at an output somewhat greater than the point at which variable unit costs turn up.* Since total unit costs equal the sums of fixed unit costs and variable unit costs at each level of production, they vary with variations in these component parts. We have already noted that the fixed unit cost component of total unit cost falls continuously with expanding output. (The overhead cost per Ford, for example, is less the greater the output of Fords.) So far as this component is concerned we should expect total unit costs to become smaller and smaller the greater the total product of a concern. Variable unit costs also fall at first, and through production ranges in which this is the case total unit costs must be falling. (If labor and raw material costs as well as overhead costs per Ford are falling with expansions in production the total cost per car must be falling.) But when variable unit costs start rising what may we expect? Will total unit costs rise also? The answer will depend on whether or not fixed unit costs are declining rapidly enough to more than counteract the effect of the rising variable unit costs. (It depends on whether the cost reduction from spreading overhead costs over more cars is more or less important than the increasing labor cost per Ford when production is speeded up.) If the drop in fixed unit cost is the greater, total unit costs will continue to fall. But fixed unit costs decline more and more slowly, while variable unit costs rise more and more rapidly. This means that eventually we must come to a point at which the rise in variable unit cost is greater than the drop in fixed unit cost. It is when, and only when, this point is reached

that the total unit cost (which is the sum of the other two) will turn upward. In the example of Table 7—1 variable unit costs start climbing with the employment of the eighth unit of input. At first they increase by amounts smaller than the decreases in fixed unit costs, but when the tenth agent is employed the drop in fixed unit costs (1 cent) is less than the rise in variable unit costs (2 cents) and the total unit costs therefore start up. The minimum total unit cost is at an output of 3,114 units, employing 9 variable agents.

The "least-cost-combination" is that combination of variable agents with given fixed agents which results in the lowest ("least") total unit cost. Roughly, it is that combination of shoe workers, leather, eyelets and rubber soles which when employed in a given shoe manufacturing plant will produce shoes at the lowest *total* cost per pair. This position is reached in the hypothetical firm of Table 7—1 with 9 units of input yielding an output of 3,114. Total unit cost in this least-cost-combination is \$0.61. Any other set of fixed agents would result in some different least-cost-combination.

Short-run output-cost curves

Some people may more easily visualize costs per unit of output by the use of graphs. Along the vertical axis are marked off per unit costs; along the horizontal axis are total outputs. Then the data on total unit costs, variable unit costs, and fixed unit costs for each total output may

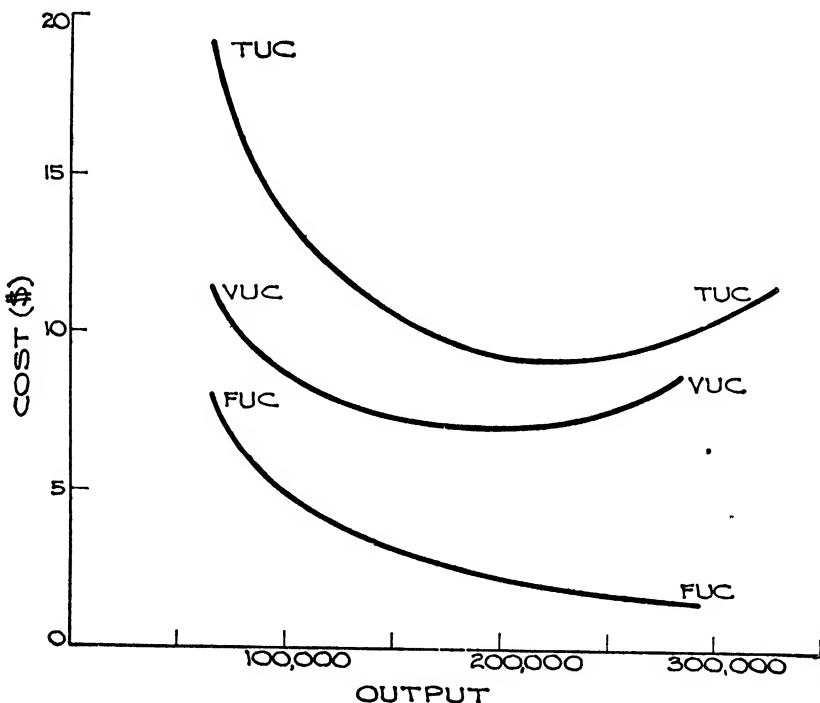


Fig. 7—1. Cost-output data for a hypothetical pencil manufacturer.

be plotted. If we plotted these data from Table 7—1 we should have 15 points for each type of costs. But if there were many hundreds of different points they would fall so close together that they would make a continuous curve. A simplified model of these cost-output relations can therefore be sketched by drawing smooth curves with the same essential characteristics as appear in Table 7—1. This is what we have done in Figure 7—1. The curves *TUC*, *VUC*, and *FUC* represent total unit cost, variable unit cost, and fixed unit cost curves respectively for a hypothetical firm. The drawing of smooth curves involves the assumption that each point on the curve refers to some cost-output relation. This implies the possibility of varying outputs by one unit (or even fraction of a unit) at a time, an unreal but a useful abstract simplification.

Let us suppose that these cost curves represent data for a concern manufacturing "ever-sharp" pencils. How could we interpret them? We might ask what the various cost figures are when the output of pencils is 200,000. The point on the total unit cost curve (*TUC*) corresponding to an output of 200,000 pencils tells us the total unit cost of producing the output of 200,000; it is about 9.3 cents. Similarly, from the fixed unit cost curve (*FUC*) we find the fixed unit cost of producing the output of 200,000; and from the variable unit cost curve (*VUC*) we find the variable unit cost at an output of 200,000. Fixed unit cost at this output level seems to be about 2.3 cents, and variable unit cost is about 7 cents. At any particular output, the cost indicated by the total unit cost curve is the sum of the figures shown as variable unit cost and as fixed unit cost. At the output of 200,000 pencils the fixed unit cost of 2.3 cents plus variable unit cost of 7 cents add up to the total unit cost figure of 9.3 cents.

The decision to operate or to shut down in the short run

By selling various outputs, a firm can get different amounts of gross income, designated in economics as "total revenue." If there is any input-output adjustment at which it is possible to obtain a total revenue greater than the associated total cost, profits are possible; otherwise the best the firm can do is to minimize losses. When returns more than cover both fixed and variable costs, so that there are profits, it is obviously reasonable to continue production.

When losses are inevitable a major decision must be made: Will it be better to continue to operate or to shut down? In the short run, fixed costs have already been incurred; there is no choice open to the entrepreneur with regard to these costs. They may be met, they may not. There is no guarantee, for example, that bondholders will get back the funds they loaned to cover investments in buildings and machines. If the firm should shut down they would get no return at all; if it operates they may get something provided that revenues are more than enough to cover the variable costs incurred (for labor and other agents) in operating. These variable costs are still optional, the entrepreneur may decide to incur them or not; he will certainly not do so if he expects that revenues will not cover the operating costs of the enterprise. *Thus the answer to*

the question whether a concern will continue operating in the short run even though it is not covering all costs, will turn on whether it is possible to cover at least variable costs. Any return over and above variable costs can be applied toward fixed costs, and something is better than nothing. Failure to cover even variable costs would simply get the firm in deeper and deeper.

A Long-Run View of Costs

The long run has been defined as a planning period long enough to permit consideration or reconsideration of the employment of all the agents in an enterprise, of the incurring of all kinds of costs in that enterprise. Long-run planning considers all possible alternatives in the combining of productive agents, whereas in short-run planning variations are possible only within the range permitted by the existing "fixed" productive agents. In the long run, all costs are variable, and long-run adjustments include the entry of new firms into an industry or the withdrawal of firms from it.

The "optimum scale of enterprise"

In the making of long-run production plans, the entrepreneur has before him a wide range of alternatives which include the construction of various kinds and sizes of plants and numbers of plants as well as the operation of any given plant with different amounts of labor and raw materials. By "scale" of establishment or of enterprise is meant the amount of investment in relatively fixed agents. For each scale of plant or enterprise there will be some level of operations at which total unit cost will be minimized; in the hypothetical example of Table 7—1 this "least-cost-combination" of variable agents with particular fixed agents gave a total unit cost of \$.61. For some sets of relatively fixed agents the least-cost-combination will give a lower total unit cost than for others. *That scale of enterprise at which the least-cost-combination is lowest is called the "optimum" scale.*

This idea may be clarified by an illustrative case, say an entrepreneur considering the manufacture of airplanes. He could build a very small plant, turning out a small number of planes, each plane almost custom-made. By combining different quantities of labor and material with that plant he could produce different quantities of planes at different total costs per plane (total unit costs). There would be some least-cost-combination, given this basic equipment, which would minimize his total unit costs. But this type of production will probably at best be very costly. If he decides to build a larger plant there will probably be some combination of productive agents that will bring out planes at a lower cost per plane, a lower total unit cost, than would be possible in the smaller scale of enterprise. On the other hand, the plant might be "too big" to give the lowest possible production costs per plane. It might become unwieldy and the least-cost-combination might be at a total unit cost higher than in a plant of more moderate size. That scale of plant

or enterprise which makes possible the very lowest total production costs per plane is the "optimum" scale.

"Economies" and "diseconomies" of scale of enterprise

"Economies of scale of enterprise" are said to exist so long as with larger and larger investments in relatively fixed agents the "least-cost-combination" involves ever lower total unit costs. So long as further increases in the investment in relatively fixed agents in a firm would bring "economies of scale of enterprise" the optimum scale has not yet been reached.

"Diseconomies of scale of enterprise" exist when with larger and larger investments in relatively fixed agents the "least-cost-combination" involves ever higher total unit costs. If investment in relatively fixed agents in a firm were so great that diseconomies of scale of enterprise had set in, that firm would be larger than the optimum.

Many, though not all, of the factors discussed in Chapter 5 as influencing the size of business enterprise are factors causing economies or diseconomies of scale of enterprises under different sets of conditions. Technological factors, for example, largely explain the economies of scale of enterprise as between extremely small and moderate sized firms; and probably management problems largely explain the diseconomies which may set in if enterprises grow very large.

Long-run output-cost curves of the firm

The relation between costs and scale of enterprise may be illustrated by drawing a set of total unit cost curves, each curve applying to a different

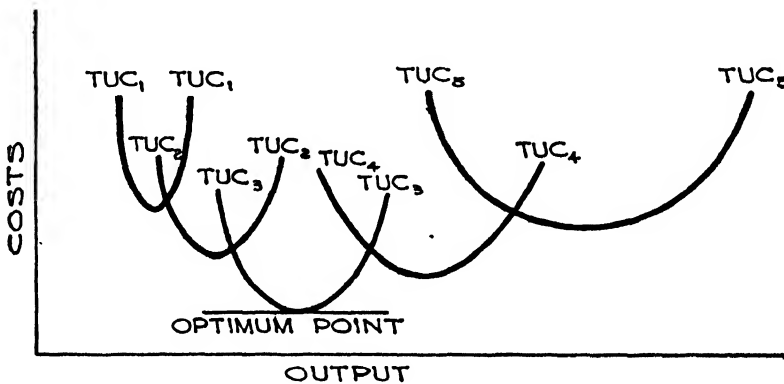


Fig. 7—2. Long-run cost opportunities of a hypothetical firm.

set of given fixed agents. Such a set of total unit cost curves is presented in Figure 7—2. The optimum scale of enterprise is clearly that represented by the cost curve TUC_3 . At any larger or any smaller scales it would be impossible to attain as low a level of total unit costs. In the range from TUC_1 to TUC_3 there are "economies of scale enterprise," the successive least-cost-combinations are at lower and lower levels. In the

range from TUC_1 to TUC_2 , there are "diseconomies of scale of enterprise," the successive least-cost-combinations are at higher and higher levels. In some cases there may be several scales of enterprise that make possible the same minimization of costs; these would then all be "optimal" scales.

Divergences from optimum scale

A statistical survey of any particular industry would probably reveal a wide range in scales of the several firms. These differences can be only partly explained by variations in special conditions of location, management talent, and so on, which would cause the optimum scales to be different. These firms also vary in scale because for many reasons some of them have not arrived at the optimum position and some have expanded beyond it. In fact it is likely that at any point in time the scale of the majority of firms in an industry will be approximately what was considered to be an optimum at some time in the recent past when they were planned. New firms will be planned to reach the optimum as it is today.

The reasons why in practice few firms are at the optimum scale at any given time are mainly: (1) errors of judgment, (2) lags in growth and shrinkage of the scale of the firm, (3) limited demand for the product relative to optimum scale, (4) attempts to control the industry, which may make it to the advantage of a firm to expand beyond the optimum scale, and (5) the special interest of insiders able to control an enterprise and gain from actions which may incidentally have the effect of diminishing the profits of the concern.

1. Errors of judgment are inevitable in the making of complex decisions such as are involved in the initiating of new enterprises. Even though a textile manufacturer, for example, may base his decisions on a careful evaluation of experiences of firms in the textile industry in an effort to determine what would be an optimum scale, he may still construct plant larger or smaller than the optimum. His estimates include allowances for changes in technology, estimates of the wages he will have to pay to get labor by the time his plant is ready to operate, evaluation of his own qualifications as a personnel man in a small intimate organization or an executive in a more extensive concern, and many other things. That all of these questions will be answered with precision and accuracy is most unlikely.

2. There are many reasons for lags in adjustments of scale of enterprise; most of these reasons are part of the growth cycle of a firm and of the rigidity of many fixed costs. It is rarely possible to finance at the start an enterprise of optimum scale if investment required is large; risks are too great. In many cases, therefore, the enterprise will grow as it is successful and finally attain an optimum level of efficiency. If the optimum scale changes there will again be a lag before it will be attained, a lag that will be much greater when the optimum shifts downward than when it moves up.

3. There are many situations in which the demand for a product is not

great enough to justify expansion of a firm to the optimum scale. Many of the local public utilities, such as light and power plants, are clear examples. If a firm could not sell the output it could efficiently produce at its optimum scale, it would be absurd to expand to the optimum level. Optimum is only the *lowest* cost scale of enterprise, it is not necessarily the "best" in all market situations. It may be possible to produce smaller outputs at a lower per unit cost in a small than in a large scale enterprise, despite the fact that if demand for the product were great enough costs could be lowered by enlarging the enterprise and its productive capacity.

4. Firms may be overexpanded in order to attain a position of dominance in an industry. Concerns in such positions are able to hold up the prices of the things they sell and to depress the prices of some of the things they buy. Thus the big tobacco companies have been able because of their dominant position to hold up the prices of pipe tobacco and cigarets and to push down the prices paid to tobacco farmers. These advantages in buying and selling may more than compensate for the loss of production efficiency in expanding beyond an optimum scale.

5. Quite apart from profits, even at the sacrifice of some profits, men with dreams of industrial empire will expand their organizations way beyond reason in an orgy of bigness. Financiers may see in some expansions and combinations, as the big utility holding companies, an opportunity to make big gains for themselves at the expense of the stockholders. These departures from policies directed toward the maximizing of the profits of the firm are important when they occur, and will be given some attention later on. For the present, however, all departures from maximizing profits of the concern as the goal will be ignored.

Total costs and the entry or withdrawal of firms

Wherever opportunities for profits appear, new firms will be established provided that there is no restriction on the entry of new firms into the industry. Profits are simply the excess of total revenues over total costs. Whether or not total costs can be covered by an enterprise will determine whether a new firm will be established, and if the firm is already in existence this consideration will determine its continuance or withdrawal when it comes time to replace equipment.

The total costs that are important from this point of view, however, are not the costs which were incurred in the past, not these "original costs" for they are, in the colloquial phrase, "past history." The thing which is important is the various cost possibilities involved in long-run planning for the present and future. It does not matter to the entrepreneur contemplating entry to an industry in 1942 that in 1938 machines cost \$2,000 each; he is interested in what they cost now and in what revenues he may anticipate in the future from their use. It does not matter whether existing firms are making profits in relation to the "original costs" which they incurred in past investments; it matters only that he should be able to cover the current and future costs which he will incur if he establishes a

new enterprise or reinvests in equipment in an old one. If there is some scale of enterprise, some combination of productive agents that will make possible the obtaining of revenues greater than total costs as anticipated now, this will be an attractive undertaking; otherwise new firms will not enter and old firms will not replace their equipment when it wears out.

CHAPTER 8

The Firm in the Market

A BUSINESS enterprise faces in two directions. On the one hand it buys raw materials, labor services, land or machines; on the other it sells a product or a service to customers who may be either ultimate consumers or other firms. The ways in which an entrepreneur seeking to maximize profits will behave will depend on the various situations he faces in these two market areas. Before studying the many variations and complexities of market adjustments as they appear in the world around us, we shall fortify ourselves by developing a few simplified models with which we can compare these concrete phenomena. Two parallel sets of models may be explored, those characterizing different types of seller positions, and those characterizing different types of buyer positions.

Classification of Types of Market Positions

Pure competition in selling and buying

A wheat farmer has many competitors offering exactly the same commodity for sale. If he were to charge a higher price than they, he could sell no wheat at all; on the other hand at or below market price he can sell all he wishes. Individually, his influence on price is infinitesimal, and he must regard price as determined by forces beyond his control.

A consumer-buyer of bread is one among many other buyers of the same commodity. Sellers ordinarily do not care who it is that buys their bread, but only that someone buys it. A consumer could get no bread at all if he were unwilling to pay the market price for it; on the other hand at or above the market price he can get all he wishes. Individually his influence on price is infinitesimal, and like the wheat farmer he must regard price as determined by forces beyond his control.

A wheat farmer is selling in a "purely competitive" market. A consumer-buyer of bread is buying in a "purely competitive" market. Pure competition is thus characterized by two things: ¹

¹It is also assumed that each buyer and seller has full knowledge of prices at which goods or services are being bought and sold by other traders in the same market.

1. There are many competitors (whether buyers or sellers).
2. Each competitor offers or seeks exactly the same thing as do the others. There is nothing to distinguish one from another.

When there are many sellers or many buyers of the same product, no one of them will be in a position *acting alone* to exert a significant effect on price. Each will act as if market price were determined by outside factors, and will simply adjust his actions to it. Yet the determination of that market price will depend on the actions of *all* the hundreds or thousands of individuals involved in the competition.

Simple monopoly in selling; simple monopsony in buying

Quite different is the case of the Aluminum Company of America, the United Shoe Machinery Company, or the local power and light company in any small town. Until recently there was only one producer of aluminum ingots; at the present time the production of some kinds of shoe machinery is controlled by one firm under exclusive patents; most local production of electricity is usually concentrated in one enterprise. Here are striking examples, not of pure competition but of monopoly. The monopoly in no case is complete: other metals may be used instead of aluminum for some purposes, but in the light-metal field there is at present little competition from substitutes; there are other ways of making shoes than by the use of the machines of the United Shoe Machinery Company, but they are very inefficient alternatives; the local power company is in a strong monopoly position as the sole provider of electricity, since the substitution of candles and oil lamps by consumers who rebel at high electric rates are only very remote possibilities.

The Aluminum Company of America, the United Shoe Machinery Co., the local public utility, each of these may be termed a "simple monopoly." No one else offers the same goods or services. Rivals are so remote as to be safely ignored in most of the policy decisions aimed at the maximizing of profits. Were its policies to remain unregulated by governmental agencies, the Aluminum Co. of America could set whatever price it might wish; so could the United Shoe Machinery Co., and the local power and light company. The only limitation would be the behavior of buyers of the several products; in general at higher prices it will not be possible to sell as large an output, and conversely larger outputs can be sold only at lower prices. Many simple monopolies are left pretty much to formulate and carry out their own price and output policies; but local power and light companies are usually strictly supervised by governmental commissions which prevent them from boosting price regardless of costs and public welfare. "Simple monopoly" is thus characterized by two things:

1. There is only one seller of the good or service.
2. Rivalry from the producers of substitutes is so remote as to be insignificant.

Under these circumstances the simple monopolist is in a position to determine price.

When there is only one buyer of a good or service, an analogous situation appears. Examples are probably most common in the buying of labor services, and striking cases emerge when a large company sets up an entire new town in an isolated area, and ships in its own labor. It is the only buyer of labor in that town, and is therefore in a strong position in the labor market. Because of this strong position it can determine wages within a wide range. Laborers seeking alternative outlets for their services would have to incur high costs of moving; moreover they are likely to be uninformed about opportunities elsewhere. Within the town they are at a disadvantage because there are no other buyers of labor in close competition with the single firm. This control over the buying side of a market is commonly designated "simple monopsony" to distinguish it from its counterpart "simple monopoly" on the selling side. The essential characteristics of "simple monopsony" are:

1. There is only one buyer of the good or service.
2. Rivalry from buyers who offer substitutive outlets is so remote as to be insignificant.

Under these circumstances the buyer is in a position to determine the price he pays for the goods or services he buys.

Monopolistic competition in selling; monopsonistic competition in buying ²

Frequently the seller of a product that is somewhat different from any other product is nevertheless in close rivalry with producers of substitutes. If his rivals are many, he will be in a position very much like the purely competitive seller; but since his product is somewhat differentiated from others, he will have some slight degree of direct control over the price he individually can charge. He will, however, have very little part in determining the general level at which his product and the close substitutes for it will be priced. Here is a situation which appears frequently in retail markets where different brands, let us say of canned peas, are sold in close competition although each brand may command a slightly different price; the same monopolistic-competitive situation arises where the location of a store or the personality of the sales clerks enables the store manager to vary prices within a narrow range from the general market level. This type of market situation is frequently described as "monopolistic competition": there are elements of both simple monopoly and pure competition in it. Hence the major characteristics of "monopolistic competition" are:

² The definition of monopolistic competition used here is narrower than the usual connotation. Instead of covering all cases of slight product differentiation it excludes all oligopoly, retaining only the case of slightly differentiated sellers where rivals are numerous and in equal degrees of rivalry, so that reactions of rivals are not taken into account.

1. Each seller is slightly differentiated in his offerings from the others. (Differences may be in conditions of sale as well as in the composition of the good itself.)

2. There are many rival sellers.

Analogous is the situation when rival buyers provide closely substitutive outlets to those seeking to sell them goods or services. Mills in adjacent towns compete with each other for the services of weavers and spinners who may prefer one town to another. Factories buying equipment differ in their credit ratings, in the promptness with which they pay, in the likelihood that they will be continuous customers. When buyers are thus slightly different they may make purchases of the same goods or services at slightly different prices. The possible control over prices that any one of these buyers may alone exert is, however, very limited. The general level of price is beyond his control, since he is only one among many buyers who are in close rivalry with each other. This situation is logically described as "monopsonistic competition"; it is a hybrid between pure competition in buying and simple monopsony. The characteristics of "monopsonistic competition" are:

1. Each buyer provides an outlet slightly differentiated from the others.
2. There are many rival buyers.

Oligopoly in selling; oligopsony in buying

Markets in which there are only a few sellers or buyers are also numerous. Degrees of concentration of production in a few firms was a topic of discussion in Chapter 5, where we saw that a high degree of concentration occurred most commonly in transportation and other "utility" services and in certain branches of manufacturing. When there are only a few firms selling or buying a good, each is aware of his individual rivals and takes them specifically into account. Acting alone he may exert a direct and important influence on price, but he will not determine his policies without taking his rivals' possible action into account. He may even join with them in formal agreements involving group determination of policies.

The significance of this interdependence in the policies of each of a small number of firms is easily illustrated. Suppose that there are just a few manufacturers selling washing machines. Firms in the industry would like to get high prices for their products, and through formal or informal agreements standards of price and quality are set up. Each company may produce a \$59 machine, a \$69 machine and a \$79 machine; each firm provides certain features and gadgets on the \$79 machines that are not included on the \$59 machine. Now suppose one of these firms which we may call the "Smooth Service Washing Machine Co.," were to try offering its \$69 type of machine for \$62 in order to attract more customers. Entrepreneurs of the other companies might get "plenty mad," and they would probably cut prices also, "going after" the "Smooth Service Washing Machine Co." The initial cutting of price to \$62 will hardly

be tried without a careful weighing of the resulting actions to be expected of rivals. No single firm selling in an industry of only a few enterprises could afford to ignore rivals' reactions. This is the basic characteristic of oligopoly.

The situation where there are just a few rival buyers is analogous to the case of oligopoly among sellers. Suppose that there were just four shoe factories in the St. Louis region. These factories would sell shoes in national markets in competition with factories located in New England and other regions; they are not oligopolists in the selling of shoes. But they are *oligopsonists* in the buying of services of skilled shoe cutters. Since (by assumption) there are only four firms in the region each firm will take into account the reaction of rivals before it decides what its wage policies in the hiring of shoe cutters will be. Each firm will hesitate to bid up wages in an effort to get more workers, since this procedure merely precipitates the bidding up of wages in rival firms. *No single firm buying in markets where there are only a few rivals could afford to ignore the reactions of these rivals to policies that it might initiate. This is the basic characteristic of oligopsony.*

Meat packing is a good illustration of an industry that is made up of only a few firms in both the selling and the buying sides of the markets. Four big packing companies buy practically all the animals to be slaughtered. The same four companies constitute the only significant sellers of the processed meats. This concentration constitutes "oligopoly" on the selling side, and it constitutes "oligopsony" on the buying side so far as purchases of animals is concerned. It does not, however, constitute oligopsony in the buying of the services of common labor since there are many other outlets for such labor services besides employment in meat packing firms.

Oligopoly and oligopsony may appear in markets in which the sellers or buyers respectively are just alike or in which they are slightly differentiated. When the offerings of the oligopolistic sellers (and the conditions of sale) are identical, we speak of "homogeneous oligopoly"; when there are differences between them it is "heterogeneous oligopoly." Most concrete situations are heterogeneous rather than homogeneous.

While the distinction between homogeneity and heterogeneity is important, the essential characteristic of oligopolistic and oligopsonistic markets is the really basic thing. For in these markets are only a few rivals; as a consequence they take each other's reactions specifically into account, whether tacitly in the separate formulation of policies, or jointly in agreed group policy determination. In this way they may quite successfully control price over considerable periods of time.

Freedom and Restrictions of Entry of Firms in the Long Run

The types of situations which we have just classified refer to positions of firms already in a market at any given time. They have not involved

the consideration of opportunity or lack of opportunity for entrepreneurs to establish firms in an industry or to withdraw when they see fit. Entrepreneurial policies concerning prices, inputs, and outputs are based, however, not only on the short-run considerations of the status quo among existing firms, but also on a consideration of possible competition from new enterprises. If there are no barriers to the entry of new firms, profits will tend to attract new firms. If there are strong barriers newcomers may be quite effectively excluded.

If profits can be made growing soybeans, more farmers will begin to grow soybeans; that is, there will be new firms in the soybean-growing industry. The same thing will happen if there are profits obtainable in growing cotton, or corn, or potatoes, or any of a large number of agricultural products. There is nothing to prevent a new entrepreneur from joining the ranks in these industries: large funds for investment are not required; there are no patents which exclude outsiders; there are so many firms already in the industry that there is no possibility of successful collusive action to keep out newcomers. These industries are characterized by "freedom of entry" in a complete sense, that is, freedom to duplicate exactly the positions of firms already established in the industry.

The situation in retailing is somewhat similar, though no retailer exactly duplicates the offerings of others. Conditions in manufacturing are extremely diverse. Many types of consumers' goods, such as shoes, dresses, socks, are manufactured by a large number of enterprises, and new firms could easily appear producing close substitutes. Wherever there is advertising and where trade marks have been obtained, it is impossible for a new firm to duplicate exactly the product of existing firms; nevertheless, there is essentially still a large measure of "freedom of entry into the production area" in which the products of the various firms are very close substitutes.

The farmers in our illustration are purely competitive sellers; the producers of retailing services and of shoes, dresses, and socks are in a situation characterized by larger degrees of monopolistic competition. The conditions which lead an industry to be purely competitive or monopolistically-competitive are also conditions in which the entry of new firms is likely to be either completely or relatively free. There is no single dominant firm or group of firms holding exclusive controls.

If we look at the Aluminum Co. of America, or at the major petroleum companies or the meat packers, we see a very different picture. Patents on essential processes are used to keep out new enterprises. If necessary, small new concerns are bought up by the established firms to prevent their development as effective competitors. In some cases unfair competitive tactics are used: cut-throat price wars, for example, or unfair manipulation of favored positions such as the ownership of the oil pipe lines by the major petroleum companies. Where the optimum scale of enterprise is large relative to the size of the market, either oligopoly or monopoly will inevitably arise. This would not in itself necessarily involve restrictions on entry, but it does provide the opportunity for the development

or techniques of restriction of entry which in turn support the position of dominance of the single firm or small group of firms in the industry. Many simple monopolies exist, or have existed, not because of economies of scale of enterprise but because a single entrepreneur has managed to get into a position from which he could exclude all rivals.

CHAPTER 9

Economic Theory and Social Policy

ON THE street corner Tom, John, and Bill are arguing about the policies of Congress and the President of the United States, of the bosses of the local factory, and of the trade union leaders. At the club Mr. Rockefeller, Mr. Morgan, and Mr. du Pont are discussing the affairs of the nation, the affairs of the businesses in which they have special interests, and the national and local affairs of labor. In the many homes of America these same issues are being argued. Some of those participating in these discussions are highly trained; others are everyday lay citizens. They are criticizing and judging both public and private policy. They are suggesting to each other remedies of varying degrees of merit and insight or lack of insight. None of them knows all the answers. Some know more than others. And they do not all want the same things anyway, though there are many things they seek in common. What can economic analysis contribute to the solution of these many questions? What are its limitations? The purpose of this book is to find paths through the maze of economic questions that arise today, paths which will guide us both in our actions as private individuals and as citizens of a large community. If we are to follow these paths, it is well at once to understand how they are marked and where they are likely to go; to recognize that they cover only limited phases of the total complex of social problems. It is that orientation that is the purpose of this chapter. Before examining the role of economic analysis in policy-making, however, we shall give brief attention to some of the ends toward which social policy in the economic sphere might possibly be directed.

Some Frequently Accepted "Ends" of Social Policy

The problem of ascertaining the ends, or goals, of social policy is a very complex one. How, in a democratic society, can the attitudes of the majority of people be discovered? It is not enough to take a poll of opinions. Many attitudes are only half-conscious, and most people are only partially articulate. And apparently subtle variations in attitudes are sometimes very significant in revealing underlying differences. Moreover, human beings are neither consistent nor fully informed. Each individual has ideas which are mutually incompatible. Conflicts will be even greater between different individuals.

To define accurately the social ends of "the American people" is obviously impossible. There is no such thing. But there are a number of ends which have been formulated by people interested in that vague thing called "social welfare." Some of these have characterized popular thinking for years; some seem to be increasing in importance; others are receiving less attention than formerly. Among these many ends are some in the attaining of which economic activity is more directly involved than in others. In order to evaluate any policy there must be some end or ends to which discussion is related. We are therefore listing here five social ends that are relevant in the evaluation of the functioning of the economic system and of policy relating to that system. We have sought to include those which would be regarded as desirable by either popular masses or significant groups of thinkers in America today. As set up, there is considerable overlapping and conflict between them. They are:

1. Allocation of resources in accordance with consumer preferences.
2. Freedom in the choice of a job or business.
3. Equity in the distribution of incomes.
4. Progress in raising planes of living; elimination of waste.
5. Stability and security.

Allocation of resources in accordance with consumer preferences

Nineteenth-century individualism placed great emphasis on the free allocation of resources in accordance with consumer wishes. It was assumed that a private enterprise economy served very well to bring about this result. Consumers express their preferences for goods and services in the market by spending the money incomes they receive in return for their contributions to production. It was assumed that producers responded appropriately to the consumer preferences thus expressed. It was also assumed that individuals knew their own interests better than did anyone else and that their dollar expenditures were accurate indexes of their wants. Today some of these assumptions are seriously challenged; but the social goal of realizing a situation in which consumer wants are effectively expressed and responded to by producers remains an important one.

To realize complete response to consumer wishes in the allocation of resources would be not only impossible, but would to some extent result in conditions inconsistent with other goals. The focus on consumers' positions ignores questions of equity in the distribution of income among the members of a society. It takes any existing income distribution for granted; people with larger incomes have more money to spend and hence exert a greater influence on the allocation of resources. Such allocation is contrary at some points to the realizing of producer preferences. And a program giving maximum voice to individual consumers may result also in a high degree of instability and insecurity. With shifting emphasis toward these "other" ends, the end of maximizing consumer influence loses some of its previous importance. At the same time,

because economic developments make its attainment even more difficult than previously, many people are emphasizing the need for new study in an effort to preserve and increase the consumer's voice in the use of resources while taking into account other goals as well.

There has always been a direct conflict between the idea on the one hand that consumers should act as free agents influencing the allocation of resources and the idea on the other hand that consumption of certain things such as opium should be discouraged or forbidden, and consumption of other things such as education and vaccination should be especially encouraged or even required. Insofar as attempts to control directly the consumption of particular goods or services have appeared in our society, they constitute exceptions to the general goal of free consumer choice and determination of resource allocation.

Freedom in the choice of a job or business

Quite as important as consumer freedom in the philosophy of individualism is the freedom for each member of the society to work at the job of his choice, to become a mechanic or carpenter or a teacher or lawyer or doctor or businessman or farmer according to his tastes and abilities. This goal is consistent with that of giving consumers freedom to direct the use of resources in accord with their preferences only insofar as the sole aim in choosing a job is the maximizing of income. The labor resources which reside in each individual member of the society are bid for by consumers through the link of business enterprises; and men and women seeking to offer their services in return for money incomes will have the free choice of doing one thing or another depending on how they individually balance the income opportunities and the pleasantness of the respective occupations in their own minds. This freedom of the individual to choose what kind of role he will play in the production of goods and services is one of the most cherished of individual freedoms. That there are many factors which interfere with its attainment today goes without saying. Many of the restrictions on freedom in the choice of job or business grow out of the development of powerful economic groups such as "big business" oligopoly, and exclusive labor unions. Other restrictions are evident, for example, in different opportunities for financing the training necessary for some kinds of work, and in customs which work to the prejudice of particular large groups such as Negroes and Mexicans in the United States.

Equity in the distribution of income

"Equity in the distribution of income" is a pleasant phrase which can be comfortably accepted in a vague way by everyone; by the same token it is a meaningless phrase unless its content is further elaborated. There are three aspects of the idea of equity of income distribution which have had at least wide verbal acceptance in the social thought of Americans, though by no means universal acceptance.

1. Associated with the rugged individualism of nineteenth-century America was the attitude that each man should receive in income the amount he contributed to the total production of the society. This idea has been given more precise form by the intensive theoretical analysis of economic relationships. Such analysis has frequently been used to justify the *status quo*; incomes received because of the productive contributions of owned material or productive resources such as land or machines have not been distinguished from income resulting from the contribution of labor. This elaboration will be considered later. For the present we shall be content with the rough statement that equity in distribution is sometimes regarded as "distribution of income according to economic contribution." But since (a) some people in fact have many more occupational opportunities than do others, (b) some people are more capable than others, and (c) some people have much more income from property than do others, the complete realization of this goal, even if it were possible, would leave a few at near-starvation level and others with very high incomes. The philosophy of individualism therefore requires as a corollary equality of opportunity before it is thoroughly acceptable to the humanitarian. Universal public education is an expression of this attitude. Even if such equality of opportunity were attainable income distribution would still be very unequal.

2. It is generally accepted as socially desirable that everyone should have at least a sufficiency of goods and services to live decently and maintain health, that society should provide at least this minimum for everyone. Disagreement arises as to (a) what level is a minimum, and (b) what should be done about loafers and ne'er-do-wells. The first question is partly a matter of individual opinion, partly something for nutritionists and other technically and professionally trained people to decide. The second problem is a very subtle one. When a man is able to work, and has an opportunity to work but does not, how should society handle the problem? The answer given will depend in part on how his behavior is explained.

3. Probably most Americans regard the present distribution of income through the private enterprise economy as inequitable because it is so unequal. Guaranteeing at least a minimum to all involves a leveling up of incomes at the bottom, but this is generally not regarded as enough. The filling in of low incomes should, according to this view, be accompanied by a paring down of the high incomes. Many aspects of social policy, especially tax policy, sometimes take this attitude into account.

Progress in raising planes of living

The American heritage has been a dynamic one; "pioneering and progress" are deeply imbedded in American ideology. The rugged individualism of the past two centuries and the early emphasis on progress were related first to the conquering of a pioneer country, then to the developing of ever new techniques that would increase productive power.

In the 1930's "progress" seemed to be taking on a new meaning. There

were those who wondered whether technological progress had been "too rapid." People were increasingly concerned with improving the social-economic organization in order to realize the potential productivity of existing resources and technology. This improvement in economic organization seemed to be the crying need if planes of living were to be raised and maintained. These new attitudes concerning the meaning of progress were more subtle than the earlier emphasis on invention, but still they could capture the imagination as did the pioneering and the inventions of earlier generations.

World war changes this focus for a time. "Total war" means organization of the economy for war, and civilian planes of living are something to be curtailed to release resources for war production. Technological progress becomes tremendously important in the effort to find substitutes for scarce and vital products and in the effort to develop superior weapons of destruction.

Stability and security

Emphasis on stability in economic life has greatly increased with the extreme fluctuations of recent years. Price changes, changes in employment, changes in the total output of the economy—these and other changes bring with them suffering for many individuals and uncertainty and insecurity for almost all. A degree of security may be provided in social legislation guaranteeing at least a minimum of living to all. But this does not guarantee that Bill Smith will not lose his house, his job, his social status. It does not guarantee that the skill which put him in the aristocracy of labor will not be displaced by the introduction of new technologies. It does not guarantee that plans to send his children to college, or even to keep them in school through high school years will not be upset. While he will probably be able to get food and shelter of a sort in his old age, he is not guaranteed either comfort or respectability. Planning for the future may seem ridiculous to those whose futures are so uncertain that their plans may be shattered by a puff of the economic winds. In such a situation stability and security seem worth some sacrifice of the possibility of attaining higher planes of living.

This increased emphasis on stability was in fact closely associated with the shifting concepts of "progress" in the 'thirties. Progress in the organization of economic life to get a more complete return from existing resources and technology meant primarily progress toward the elimination of depression periods of severe unemployment and diminished production.

The Role of Economics

Any particular economist is of course a human being; being an economist is only one aspect of the total person. He may be many things besides an economist; and in these other roles he may join the Republican or the Socialist party; he may belong to a Quaker meeting or to the

Episcopal church; he may make speeches about going to war or not going to war, and so on. For the moment, however, we are not concerned with an economist as a total person. "The economist" is just an aspect of a person. What is his particular role?

The primary task of the economist is the understanding and explanation of the interrelations between actions in the various sectors of economic life. What difference does it make whether there are many or few sellers of a good? For example, what are the significant differences in prices and outputs resulting from the actions of the many individuals directing production and sales in agriculture, the few in the steel industry, the single corporation (until recently) in aluminum production? How will the actions of trade-unions affect employment, the wages of workers here and elsewhere, the prices paid by consumers, the goods produced? What are the effects of advertising on planes of living? How will tariffs affect the distribution of incomes and the prices and quantities of goods? What will be the repercussions of particular government action intended to restrict agricultural outputs or to maintain the prices of agricultural products? Who ultimately bears the burden of a tax on automobiles or cigarettes or land or buildings or pay rolls?

In answering questions such as these the economist makes several kinds of contributions to the policy-maker and to the critic of private or social policy in the economic sphere: (1) He assists in the formulation and clarification of social ends. (2) He assists in determining what facts will be economically relevant to the consideration of a problem. (3) He aids directly in the selection of means to the attaining of any given ends.

The economist and the clarification of social ends

Of all the difficulties faced by the policy-making executive, one of the most troublesome is that of ascertaining the social ends or aims involved.

Of course, when the policy is that of a private corporation functioning within the framework of a free enterprise economy the problem is relatively easy. It is usually safe to assume that the object is to maximize net income, although regularizing its flow may be of almost as much importance, and the prestige of the business enterprise is by no means without weight. Evaluating these ends and determining to what extent each may, if necessary, be sacrificed for the sake of the others will not normally be a source of much heated debate among those directing the policies of a firm.

When on the other hand the policy is that of some public agency the answer is not so easily arrived at. If the agency feels that it owes its just powers to the consent of the governed its measures must have the acceptance of the public. This, however, is not generally regarded as enough, since good government "should" strive to put the public's active wishes into effect. To some extent the political mechanisms of the state afford an avenue for the expression of these active wishes and not merely of the passive consent of the governed. But the time between elections is long, and the elections themselves are fought on many issues

and do not afford a clear-cut verdict as to the public's position on any one of them. The public itself is made up of many individuals, each with his own peculiar values, his own prejudices, and his own interests. Small wonder then that fact-finding in the realms of opinion is treacherous and relatively undeveloped. None the less, learning what ends people seek is a most significant type of fact-finding for public and private agencies alike.

The economist is not, as economist, a politician or voter. He does not decide what would be an ideal society. He does not tell anyone what social or private ends are in themselves desirable. The decisions as to what shall be the ends of action or policy rest with the members of the society and cannot be determined by any scientific rules.

But although he does not take sides in questions as to what the social ends "should" be, in the important task of defining and clarifying these ends the economist does play a part. There are two types of contributions he can make: (1) He may sometimes explain that means used to accomplish one end lead to results that conflict with another end, thus showing the policy-maker the necessity of making some kind of choice as to the relative importance of different ends. For example, the subsidizing of housing by government may mean either an increase in total government expenditures with heavier taxation or the curtailing of other government expenditures, say on general relief or on public parks. Or heavy taxation to re-distribute incomes may remove some of the incentives to production and result in lower planes of living. (2) By understanding the structure of economic life, he may assist in the formulation of ends which are themselves complex patterns of interrelationships. This latter is a very involved problem, the nature of which can be understood only after considerable exploration into economic analysis itself. It requires such analysis, for example, to give specific meaning to vaguely formulated goals of consumer determination of the allocation of resources in a complex economy, or of equity in income distribution based on the criterion of reward in accordance with "economic contribution."

The economist and the selection and interpretation of facts

To know enough about a problem to know what facts would be relevant in the formulation of social policy is to be very well informed. Suppose an untrained man were asked, for example, to work out a program designed to provide better nutrition for low income groups, or to prevent inflation in wartime. Would he know how to begin or what facts to gather? And when he had his facts would he know how to interpret them? Some things would seem quite obvious, but he could not go far without getting beyond his depth.

The specialist in nutrition could provide some of the relevant facts in regard to the first of these problems, facts which the economist would know little about and would be unable to interpret accurately. The sociologist may tell something of what the relevant food customs of different cultural groups are. But other facts, concerning sources of food

supply, the possibilities of expanding production of certain commodities, and the ways in which these products could be most efficiently moved through the various stages in the production process from farmer to consumer are more readily handled by the economist. It is necessary to know something about how economic processes work to know what facts to select and how to interpret them in developing a national food program intended to get certain food products to those needing them.

The role of the economist in selecting and interpreting facts relevant to a plan intended to check inflation is even more obvious. Some understanding of how the money system operates, of how money, goods, and services flow through the economy, is a prerequisite to any intelligent attack on the finding and interpretation of relevant facts.

The economist and the selection of means

The social policy maker is interested in how certain things can be accomplished. The economist can help him select means which would be appropriate to the ends he is seeking. In fact it was with various possibilities in mind as to means of accomplishing a task that the economist proceeded to find and interpret his facts. In suggesting possible means, the economist makes two kinds of contributions. He may say on the one hand that there is this and this and this way of getting the results sought. Consumers may be protected from oligopoly high prices and curtailed output, for example, by either (1) legal breaking up of oligopoly controls and enforcing of competitive action, or (2) direct government regulation of prices. On the other hand, each of these means has some further effects, some of which are regarded as "desirable," others as "undesirable." To take a very obvious example: the economist may tell the policy maker that he can get revenues for government use by taxing incomes, by sales taxes, by borrowing, by issuing paper money; he may then go on and suggest that by taxing incomes it is possible at the same time to redistribute wealth in favor of the lower income groups, whereas if public revenues are obtained by the use of a sales tax this will have the effect of placing a relatively greater burden on the low income groups. A more striking example is the economist's verdict on heavy taxation to finance public works in depression periods. Such taxation can of course be a source of public revenue which government pours back into the economy as payments to relief workers; but it defeats its own purpose as a depression measure, since the withdrawal of the tax funds in the first place is likely to exert a further depressing effect on private productive activity.

The institutional framework of economic analysis

There is no such thing as an "economic system" operating apart from the rest of the social structure. In Chapter 1 there was a brief discussion of the relation of the "economic system" to the total social system; the "economic system" was viewed as a set of interrelated actions which comprised certain aspects only of the whole of social structure. Eco-

conomic analysis is the analysis of these economic aspects of society. Any practical problem will have many aspects, some of them economic. Economic analysis can never provide complete answers to such problems. *Since it applies only to certain aspects of the whole, economic analysis should be used with an awareness of its limitations and of the surrounding social institutions within which and in relation to which economic forces operate.*

Assumptions concerning economic processes are frequently tacitly based on assumptions as to the social framework within which the economic action takes place. Many of these assumptions are unconscious, basic aspects of a culture being so taken for granted by its members that for most purposes there is no need to focus attention on them. The assumption in our society that the owner of an enterprise has the power to hire and fire is an example. On the other hand, failure to examine some of the noneconomic aspects of social processes may lead into serious error. Such error may be error of fact as to how people will actually respond to particular programs. Some farmers, for example, hold tenaciously to the methods of their grandfathers despite the lower costs of new methods. Or the error may be in the comprehending of the ends actually sought by the society and of the ways in which particular programs may interfere with the attaining of these ends because of "noneconomic" elements in the situation. A focus on economic aspects of social problems might, for example, result in overattention to the end of "giving consumers the most for their money" while ignoring the fact that this may require that workers move readily from place to place and that such mobility interferes with community ties and family life. *Failure to recognize the limitations of economic analysis when applied to problems that actually arise in a broad social system, only one aspect of which is economic, brings error and bias, often serious.*

Part II

EQUILIBRIUM OF THE FIRM

PREAMBLE

THE theory of the "equilibrium of the firm" is fundamentally a very simple matter. It just adds up to this. An entrepreneur wants to make profits as great as possible. He does whatever he can to accomplish that end. Profits are the difference between total revenue and total cost. He wants to make that difference as big as possible. If by hiring more workers and buying more raw materials he would increase his revenues more than he would increase his costs, he would clearly improve his position; this would make his profits bigger than before, or his losses less. If by hiring fewer workers and buying less raw materials he would reduce his revenues less than he would reduce his costs, he would clearly improve his position; this would make his profits bigger than before, or his losses less. He will be ill at ease so long as any improvement in his profit position can still be accomplished by such a change; he will be in equilibrium when he is accomplishing his goal of maximizing profits. In the short run it is only in his power to make short-run adjustments. Since the entrepreneur of a firm aims at this maximum profit position, the study of the "equilibrium of the firm" provides answers to two questions: How much of a product will he sell? How much of a resource will he buy? These are the first things we need to know to understand the workings of an economy such as that of the United States. Study of the "equilibrium of the firm" also provides preliminary answers to another question: Will consumers' interests be in harmony with the interests of entrepreneurs seeking to maximize profits? The answer to this question will interest you if you like to consume or if you want to make profits, or if you have some sympathy for people who like to consume and to make profits. Really to grasp the meaning of the answers to these questions is hard work.

CHAPTER 10

Preliminary Analysis of Short-run Policies of the Firm

WHATEVER the market situation of a firm, the principal purpose is generally the maximizing of profits (or the minimizing of losses). This is the goal of the small poultry farmer and the big oil magnate alike. But the small poultry farmer charges whatever the going market price of his produce may be, while the big oil magnate plays a major role in determining the price to be charged. Decisions as to inputs of productive services and outputs of product are made by both the poultry farmer and the oil magnate; but only the oil magnate considers price policy in arriving at his production decisions.

In this chapter we shall undertake a preliminary analysis of short-run input, output, and price policies of firms seeking to maximize profits. Decisions as to how much to produce and what quantities of productive services to buy are simply two aspects of the same thing. When a producer of pencils decides what his output will be he necessarily involves himself in decisions concerning purchases of labor services, wood, and paint. When he decides how many workers to hire and how much wood and paint to buy he involves himself in decisions as to how many pencils will be turned out. An examination of the output policies of a firm is thus simply one way of looking at decisions concerning the level of production in the firm. An examination of the input policies of a firm is the other way of looking at the same thing. If the pencil-maker is unable to influence the prices of what he buys or what he sells, this examination of input-output policies merely assumes a given set of prices for pencils and for the productive services used in producing them. But if a firm is in a dominant position in the selling of pencils, decisions to expand output are associated with decisions to accept a lower price for the pencils; decisions to curtail output are associated with decisions to charge a higher price for the pencils. Analogously, if the pencil-maker is in a dominant position in the buying of labor or materials, decisions to expand input are associated with decisions to pay higher prices for this labor and these raw materials in order to get more of them, and decisions to curtail input are associated with decisions to pay a lower price for labor and raw materials and take only the smaller quantity available at the lower rates.

Although the differences between oligopoly, simple monopoly, and

monopolistic competition are very important, they have one significant characteristic in common: individual firms so situated influence selling price in some degree (great or small), and price policy in selling is therefore considered in arriving at decisions concerning the level of production in the firm. They may therefore for some purposes be grouped together under the general label *monopolistic*, thus distinguishing them from the purely competitive situation in which a single seller has no influence, acting alone, over the price of what he sells.

Analogously, oligopsony, simple monopsony, and monopsonistic competition are alike in that firms so situated exert individually some degree of influence on the prices of the things they buy. Price policies in buying are therefore considered in arriving at decisions concerning the level of production in the firm. These types of buying situations may be grouped together for some purposes under the general term *monopsonistic*, thus distinguishing them from the purely competitive situation in which a single buyer has no influence, acting alone, over the price of what he buys.

The Production Policies of Purely Competitive Firms

Firms which are purely competitive in all their market relations simply buy at the existing market price and sell at the existing market price; individually they have no perceptible influence on price and their policy decisions are therefore confined to questions of levels of input and output, within an existing price framework. They will try to adjust to that level of production at which profits are greatest within the existing price conditions; and this will of course be the point at which the difference between total revenues of the firm and total cost is greatest.

The maximum profit (or minimum loss) level of production in a hypothetical purely competitive enterprise

The production policies of a purely competitive firm are easily illustrated by the use of a hypothetical example. In fact, the cost data presented for the hypothetical firm of Chapter 7 may be used for this purpose. The *physical* input-output data given might be characteristic of any type of firm; but it was assumed that the cost of a unit of input was \$100 whatever the input of the enterprise. In other words, we tacitly assumed that the firm in that example (Chapter 7) was purely competitive in the buying or hiring of variable agents of production, that whatever the input of the firm the price per unit of input would be unchanged. This price per unit of input is called "average expenditure." These cost data are reproduced here in Table 10—1A.

Some figures on possible revenues of the firm have been added to the table. Since we are assuming that this is a purely competitive seller as well as a firm in a purely competitive position in buying, the price per unit of output is independent of the production policy of the firm. It has been taken to be \$1.80. *Price per unit of output is called "average revenue"; and the total revenue obtained at each output is then the price per unit of output times the number of units of output.* Thus total

TABLE 10—1A

COST AND REVENUE DATA FOR A HYPOTHETICAL PURELY COMPETITIVE HOSE MANUFACTURER

A. Assuming: Average expenditure \$100; average revenue \$1.80

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Input</i>	<i>Output</i>	<i>Average exp. (price per unit of input)</i>	<i>Variable cost</i>	<i>Fixed cost</i>	<i>Total cost</i>	<i>Average revenue</i>	<i>Total revenue</i>	<i>Profit</i>
1	100	\$100	\$ 100	\$1,000	\$1,100	\$1.80	\$ 180	\$-920
2	350	100	200	1,000	1,200	1.80	630	-570
3	702	100	300	1,000	1,300	1.80	1,264	-36
4	1,152	100	400	1,000	1,400	1.80	2,074	674
5	1,700	100	500	1,000	1,500	1.80	3,060	1,560
6	2,190	100	600	1,000	1,600	1.80	3,942	2,342
7	2,604	100	700	1,000	1,700	1.80	4,687	2,987
8	2,908	100	800	1,000	1,800	1.80	5,234	3,434
9	3,114	100	900	1,000	1,900	1.80	5,605	3,705
10	3,240	100	1,000	1,000	2,000	1.80	5,832	3,832
11	3,300	100	1,100	1,000	2,100	1.80	5,940	3,840
12	3,350	100	1,200	1,000	2,200	1.80	6,030	3,830
13	3,395	100	1,300	1,000	2,300	1.80	6,111	3,811
14	3,435	100	1,400	1,000	2,400	1.80	6,183	3,783

revenue from 100 units of output is 100 times \$1.80, or \$180; total revenue from 350 units of output is 350 times \$1.80, or \$630; total revenue from 702 units of output is 702 times \$1.80, or \$1,264. Assuming that this firm manufactures ladies' silk hose, average revenue is the price of a pair of hose (in this case \$1.80), and total revenue is the total receipts of the manufacturer from the wholesalers or retailers buying the hose at each quantity of output (and hence of hose sold).

Profits are entered in column 9. At some outputs total revenue is greater than total cost, and profits at these outputs are positive; at other outputs total revenue is less than total cost, and profits at these outputs are negative (there are losses). The hose manufacturer will select that output which gives him the greatest positive profit; in this example it is an output of 3,300 pairs of hose, involving 11 units of input (packages of variable agents). There is no other production level at which profits would be as great as this, assuming the existing average revenue (\$1.80) obtained per pair of hose and the existing average expenditure (\$100) per unit of input.

Now let us suppose that the market price of hose were to drop (perhaps because of a sharp decrease in demand) to \$.45 a pair. Average revenue would be \$.45 and total revenue at each output would be much less than before. If costs remained unchanged, the firm would no longer be able to make a positive profit at any output, but losses (negative profits) at some outputs would be greater than at others. Table 10—1B is exactly like Table 10—1A except that the price per pair of hose is assumed to

TABLE 10—1B

COST AND REVENUE DATA FOR A HYPOTHETICAL PURELY COMPETITIVE HOSE MANUFACTURER

B. Assuming: Average expenditure \$100; average revenue \$.45

(1) <i>Input</i>	(2) <i>Output</i>	(3) <i>Average expenditure</i>	(4) <i>Variable cost</i>	(5) <i>Fixed cost</i>	(6) <i>Total cost</i>	(7) <i>Average revenue</i>	(8) <i>Total revenue</i>	(9) <i>Profit</i>
1	100	\$100	\$ 100	\$1,000	\$1,100	\$.45	\$ 45.00	\$-1,055.00
2	350	100	200	1,000	1,200	.45	157.50	-1,042.50
3	702	100	300	1,000	1,300	.45	315.90	-984.10
4	1,152	100	400	1,000	1,400	.45	518.40	-881.60
5	1,700	100	500	1,000	1,500	.45	765.00	-735.00
6	2,190	100	600	1,000	1,600	.45	985.50	-614.50
7	2,604	100	700	1,000	1,700	.45	1,171.80	-528.20
8	2,908	100	800	1,000	1,800	.45	1,308.60	-491.40
9	3,114	100	900	1,000	1,900	.45	1,401.30	-498.70
10	3,240	100	1,000	1,000	2,000	.45	1,458.00	-542.00
11	3,300	100	1,100	1,000	2,100	.45	1,485.00	-615.00
12	3,350	100	1,200	1,000	2,200	.45	1,507.50	-692.50
13	3,395	100	1,300	1,000	2,300	.45	1,557.75	-742.25
14	3,435	100	1,400	1,000	2,400	.45	1,545.75	-854.25

be \$.45 instead of \$1.80. Comparing the total revenues under these circumstances with total costs at each output level, we find that losses are minimized at an output of 2,908 units, produced with an input of 8 units. At that production level, total cost exceeds total revenue by \$491.40.

It may be that while the price of hose has remained at \$1.80 the average expenditure on a unit of input has risen to, say, \$600 (perhaps because silk has become much more costly). This situation is represented by Table 10—1C, which is exactly like Table 10—1A except for the increase in the expenditure per unit of input. Variable and total costs are consequently much higher than they were before. Again the best this manufacturer can do is to minimize his losses, which he would accomplish at a production level of 7 units of input turning out an output of 2,604 pairs of hose. In this most favorable (least unfavorable) adjustment his loss is \$513; at this production level total cost exceeds total revenue by \$513.

The decision to operate or shut down in the short run

The observant student may have noticed that the point at which profit is maximized (or loss minimized) is also the point at which the difference between total revenue and variable cost is most favorable to the hose manufacturer. This is inevitably the case, since fixed cost is the same at all outputs. Wherever the excess of total revenue over total cost is greatest, the excess of total revenue over variable cost must also be at a maximum. The more the firm can get back over and above variable costs the better off it will be.

There is an important difference between total cost and variable cost. Variable costs are the costs involved in deciding to operate or not, and at what level, in the short run. Fixed costs go on whatever the short-run decision. Therefore, it may pay to operate even if the firm cannot avoid making losses as indicated by a comparison between total costs and total revenues. It will pay provided that there is some level at which total revenue exceeds the outlay on variable agents. Such possibilities are illustrated in both Tables 10—1B and 10—1C despite the fact that firms so situated cannot avoid losses. Turning to Table 10—1B we find loss is least at an input of 8 units, producing 2,908 pairs of hose. At this level total revenue (\$1,308.60) exceeds variable cost (\$800) by \$508.60. If the firm shut down it would get back nothing at all on its fixed costs of \$1,000; by operating, it can cover the variable costs that will be incurred (\$800), and in addition get back \$508.60 toward the meeting of fixed costs. The remaining sum of \$491.40, which is not covered, is a far smaller loss than the \$1,000 of fixed costs that would be lost if the firm shut down completely.

If the price of hose were to drop yet further than in Table 10—1B, or if the average expenditure per unit of input were to rise still more than in Table 10—1C, it might be that there would be no adjustment in which the firm could even cover variable costs. In that case it would be foolish to incur such costs, and the firm would shut down.

TABLE 10—1C
COST AND REVENUE DATA FOR A HYPOTHETICAL PURELY COMPETITIVE HOSE MANUFACTURER

C. Assuming: Average expenditure \$600; average revenue \$1.80

(1) <i>Input</i>	(2) <i>Output</i>	(3) <i>Average expenditure</i>	(4) <i>Variable cost</i>	(5) <i>Fixed cost</i>	(6) <i>Total cost</i>	(7) <i>Average revenue</i>	(8) <i>Total revenue</i>	(9) <i>Profit</i>
1	100	\$600	\$ 600	\$1,000	\$1,600	\$1.80	\$ 180	\$-1,420
2	350	600	1,200	1,000	2,200	1.80	630	-1,570
3	702	600	1,800	1,000	2,800	1.80	1,264	-1,536
4	1,152	600	2,400	1,000	3,400	1.80	2,074	-1,326
5	1,700	600	3,000	1,000	4,000	1.80	3,060	-940
6	2,190	600	3,600	1,000	4,600	1.80	3,942	-658
7	2,604	600	4,200	1,000	5,200	1.80	4,687	-513
8	2,908	600	4,800	1,000	5,800	1.80	5,234	-566
9	3,114	600	5,400	1,000	6,400	1.80	5,605	-795
10	3,240	600	6,000	1,000	7,000	1.80	5,832	-1,168
11	3,300	600	6,600	1,000	7,600	1.80	5,940	-1,660
12	3,350	600	7,200	1,000	8,200	1.80	6,030	-2,170
13	3,395	600	7,800	1,000	8,800	1.80	6,111	-2,689
14	3,435	600	8,400	1,000	9,400	1.80	6,183	-3,217

Total unit costs and the position of the firm

Thus far we have omitted any consideration of the total costs to this firm *per pair* of hose, that is, of total unit costs at various levels of production.

In Table 10—2A we have repeated figures from Table 10—1A (omitting some columns to save space) and added to these figures computations of total unit costs. Two interesting facts show up in this table:

TABLE 10—2A

COST AND REVENUE DATA FOR A HYPOTHETICAL PURELY COMPETITIVE
HOSE MANUFACTURER: CONTINUED

A. Assuming: Average expenditure \$100; average revenue \$1.80

(1)	(2)	(3)	(4)	(5)	(6)
<i>Input</i>	<i>Output</i>	<i>Total cost</i>	<i>Total unit cost</i>	<i>Total revenue</i>	<i>Profit</i>
1	100	\$1,100	\$11.00	\$ 180	\$-920
2	350	1,200	3.43	630	-570
3	702	1,300	1.85	1,264	-36
4	1,152	1,400	1.22	2,074	674
5	1,700	1,500	.88	3,060	1,560
6	2,190	1,600	.73	3,942	2,342
7	2,604	1,700	.65	4,687	2,987
8	2,908	1,800	.62	5,234	3,434
9	3,114	1,900	.61	5,605	3,705
10	3,240	2,000	.62	5,832	3,832
11	3,300	2,100	.63	5,940	3,840
12	3,350	2,200	.66	6,030	3,830
13	3,395	2,300	.67	6,111	3,811
14	3,435	2,400	.70	6,183	3,783

1. At several production levels the price of hose (average revenue) exceeds the total unit cost involved in producing them. These are the levels at which total revenues exceed total costs. Such a relationship is not accidental. It inevitably follows that if the total revenue obtained from selling a certain quantity of hose exceeds the total cost of producing that quantity of hose, the average revenue per pair (price) must be greater than the total cost per pair (total unit cost). From this we may conclude that if there is any production level at which total unit cost is less than selling price the firm can make a profit.

2. The most profitable level of production is greater than that at which total unit cost is minimized. The firm will go beyond the least-cost-combination output. This may at first seem surprising, since when total unit cost is least the difference between price per pair of hose and cost per pair is greatest. But by producing more hose than this, profits are increased because a gain of \$1.17 per pair on 3,300 pairs is better than a gain of \$1.19 per pair on only 3,114.

158 PRELIMINARY ANALYSIS OF SHORT-RUN POLICIES

In Table 10—2B we have repeated figures from Table 10—1B, adding to these figures computations of total unit costs (which are of course the same as in Table 10—2A) and of variable unit costs (to be considered in the next section). Two interesting facts again show up in the data on total unit costs in this table:

TABLE 10—2B

COST AND REVENUE DATA FOR A HYPOTHETICAL PURELY COMPETITIVE HOSE MANUFACTURER: CONTINUED

B. Assuming: Average expenditure \$100; average revenue \$.45

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Input</i>	<i>Output</i>	<i>Variable cost</i>	<i>Variable unit cost</i>	<i>Total cost</i>	<i>Total unit cost</i>	<i>Total revenue</i>	<i>Profit</i>
1	100	\$.100	\$1.00	\$1,100	\$11.00	\$ 45.00	\$-1,055.00
2	350	200	.57	1,200	3.43	157.50	-1,042.50
3	702	300	.43	1,300	1.85	315.90	-984.10
4	1,152	400	.35	1,400	1.22	518.40	-881.60
5	1,700	500	.29	1,500	.88	765.00	-735.00
6	2,190	600	.274	1,600	.73	985.50	-614.50
7	2,604	700	.269	1,700	.65	1,171.80	-528.20
8	2,908	800	.275	1,800	.62	1,308.60	-491.40
9	3,114	900	.29	1,900	.61	1,401.30	-498.70
10	3,240	1,000	.31	2,000	.62	1,458.00	-542.00
11	3,300	1,100	.33	2,100	.63	1,485.00	-615.00
12	3,350	1,200	.358	2,200	.66	1,507.50	-692.50
13	3,395	1,300	.38	2,300	.67	1,557.75	-742.25
14	3,435	1,400	.41	2,400	.70	1,545.75	-854.25

1. At no production levels does the price of the hose exceed the total unit cost of producing that quantity. Moreover at no levels does total revenue exceed total cost. There is no possibility for this firm to make a profit.

2. The least unprofitable level of production is less than that at which total unit cost is minimized. The firm will stop short of the least-cost-combination when the price of hose is \$.45. This is true despite the fact that the loss per pair of hose is least when the total unit cost of producing hose is at a minimum. By producing fewer hose than this losses are diminished, since a loss of \$.17 per pair on only 2,908 pairs of hose is not as bad as a loss of \$.16 per pair on 3,114.

These examples illustrate two general facts: (1) If there are any outputs at which average revenue (selling price) exceeds total unit cost, profits can be obtained; otherwise the best the firm can do is to minimize losses. (2) Whenever price exceeds lowest total unit cost the most profitable output will be greater than or equal to the least-cost-combination output; whenever price is less than lowest total unit cost the least unprofitable output will be less than or equal to the least-cost-combination output; whenever price exactly equals lowest total unit cost the firm will operate at the least-cost-combination output, which will be the only

level at which it can break even. The student may easily satisfy himself concerning the truth of these propositions by experimenting with different assumed prices for hose and seeing what would be the situation of the hose manufacturer in each case.

Variable unit costs and the position of the firm

The relation between variable unit costs and price resembles the relation between total unit costs and price. So long as price is above the lowest variable unit cost it will be possible to obtain total revenues in excess of variable costs. Under such circumstances it would pay to continue operating, since something would be obtained above variable costs and could be applied toward the meeting of fixed costs. If price is less than the lowest variable unit cost there will be no possibility of getting back total revenues in excess of variable costs and the firm will shut down. At a price equal to lowest variable unit cost it will be a matter of indifference whether the firm continues to operate or not.

In Table 10—2B there are several production levels at which the price of hose (average revenue) exceeds the variable unit cost involved in producing them. These are the levels at which total revenues exceed variable costs. From this we may conclude that at a number of production levels this firm would be better off to operate at a loss than to shut down.

Graphic representation of profits and losses of a purely competitive firm

The profit or loss position of a purely competitive firm may be illustrated graphically. This has been done in Figures 10—1A and 10—1B. Figure 10—1A is based on Table 10—1A and Figure 10—1B on Table 10—1B.

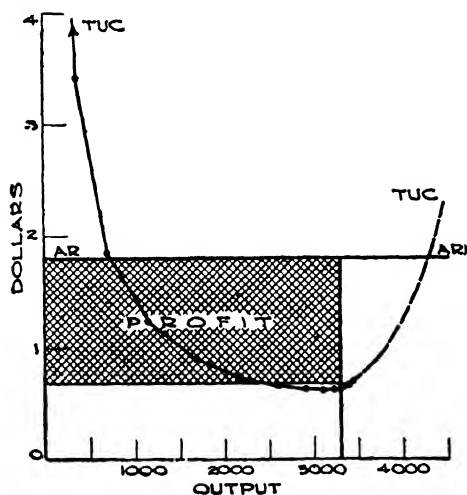


Fig. 10—1A. Profitable purely competitive firm.

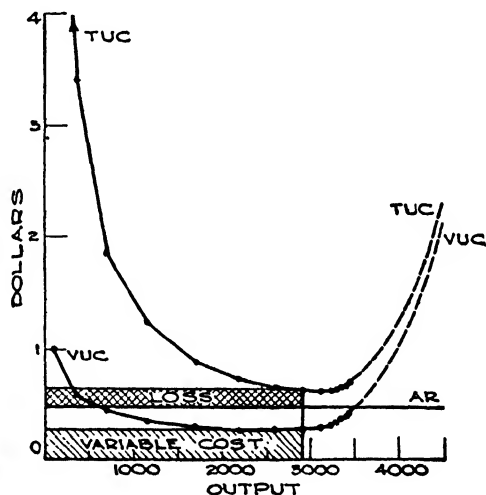


Fig. 10—1B. Purely competitive firm operating at a loss.

In Figure 10—1A the firm is making a profit. The average revenue line, drawn at \$1.80, lies well above the bottom of the total unit cost curve. There are clearly many positions in which this firm could get revenues in excess of costs. We know from Table 10—1A that its most profitable output is at 3,300 units. By drawing a vertical line at this point on the graph we form the fourth side of a rectangle the area of which represents total revenue, since one side (horizontal) is output and the adjacent side (vertical) is price per unit of output. We may read from the total unit cost curve what the total cost per unit at this output level is, and by drawing a horizontal line at this point we mark off another rectangle, the area of which equals total cost (output times total unit cost). The difference between these two areas, the shaded portion of the graph, is then profit. One side (horizontal) of the area representing profit again measures output, and the vertical height of the area represents the gain per unit of output: by multiplying the two we get a figure for profit which is the same as that found by subtracting total cost from total revenue.

Figure 10—1B is like Figure 10—1A except that the average revenue obtained for hose is assumed to be only \$.45; that is, hose are selling at \$.45 a pair. At no point does this average revenue exceed total unit cost, so the firm cannot make a profit; but at many positions it is greater than variable unit cost, so it will pay for the firm to continue operating in the short run. By turning back to Table 10—1B we may again find the best production level; in this case it is at an output of 2,908 pairs of hose. By drawing a vertical line at this point we may mark off rectangles representing total cost and total revenue just as in Figure 10—1A; but in this case total cost is greater than total revenue so that the difference between them is negative profit, that is, loss. Just as a rectangle was marked off representing total cost, so also a rectangle may be marked off representing variable cost, by drawing a horizontal line from the variable unit cost curve. This rectangle is filled in by diagonal lines. The difference between the rectangle representing total revenue and that representing variable cost is then a measure of the advantage in operating over that of shutting down. There is clearly such an advantage despite the fact that losses are inevitable.

The Policies of Firms Buying Purely Competitively and Selling Monopolistically

Production and price adjustments of a monopolistic firm

An analysis of the policies of a purely competitive firm having been pursued in detail, the case of a purely competitive buyer selling monopolistically may be discussed more briefly. Fundamentally the analysis of firms selling in noncompetitive situations is the same as that of firms in purely competitive situations; but the fact that average revenue varies with output introduces a significant difference. Tables 10—3A and 10—3B assume the same set of cost data as those of the hose manufacturer

TABLE 10—3A

COST AND REVENUE DATA FOR A HYPOTHETICAL TIE MANUFACTURER:
BUYING PURELY COMPETITIVELY; SELLING MONOPOLISTICALLY

A. Profitable situation

(1) <i>Input</i>	(2) <i>Output</i>	(3) <i>Average expenditure</i>	(4) <i>Variable cost</i>	(5) <i>Fixed cost</i>	(6) <i>Total cost</i>	(7) <i>Average revenue</i>	(8) <i>Total revenue</i>	(9) <i>Profit</i>	(10) <i>Total unit cost</i>
1	100	\$100	\$ 100	\$1,000	\$1,100	\$2.80	\$ 280.00	\$ -820.00	\$11.00
2	350	100	200	1,000	1,200	2.75	962.50	-237.50	3.43
3	702	100	300	1,000	1,300	2.65	1,860.30	560.30	1.85
4	1,152	100	400	1,000	1,400	2.50	2,880.00	1,480.00	1.22
5	1,700	100	500	1,000	1,500	2.30	3,910.00	2,410.00	.88
6	2,190	100	600	1,000	1,600	2.10	4,599.00	2,999.00	.73
7	2,604	100	700	1,000	1,700	1.90	4,947.60	3,247.60	.65
8	2,908	100	800	1,000	1,800	1.75	5,089.00	3,289.00	.62
9	3,114	100	900	1,000	1,900	1.65	5,138.10	3,238.10	.61
10	3,240	100	1,000	1,000	2,000	1.55	5,022.00	3,022.00	.62
11	3,300	100	1,100	1,000	2,100	1.45	4,785.00	2,685.00	.63
12	3,350	100	1,200	1,000	2,200	1.35	4,522.50	2,322.50	.66
13	3,395	100	1,300	1,000	2,300	1.25	4,243.75	1,943.75	.67
14	3,435	100	1,400	1,000	2,400	1.15	3,950.25	1,550.25	.70

TABLE 10—3B

COST AND REVENUE DATA FOR A HYPOTHETICAL TIE MANUFACTURER:
BUYING PURELY COMPETITIVELY; SELLING MONOPOLISTICALLY

B. Operating at a loss

(1) <i>Input</i>	(2) <i>Output</i>	(3) <i>Average expenditure</i>	(4) <i>Variable cost</i>	(5) <i>Fixed cost</i>	(6) <i>Total cost</i>	(7) <i>Average revenue</i>	(8) <i>Total revenue</i>	(9) <i>Profit</i>	(10) <i>Total unit cost</i>
1	100	\$100	\$ 100	\$1,000	\$1,100	\$.90	\$ 90.00	\$-1,010.00	\$11.00
2	350	100	200	1,000	1,200	.85	297.50	-902.50	3.43
3	702	100	300	1,000	1,300	.80	561.60	-738.40	1.85
4	1,152	100	400	1,000	1,400	.75	864.00	-536.00	1.22
5	1,700	100	500	1,000	1,500	.70	1,190.00	-310.00	.88
6	2,190	100	600	1,000	1,600	.65	1,423.50	-176.50	.73
7	2,604	100	700	1,000	1,700	.61	1,588.44	-111.56	.65
8	2,908	100	800	1,000	1,800	.57	1,657.56	-142.44	.62
9	3,114	100	900	1,000	1,900	.53	1,650.42	-249.58	.61
10	3,240	100	1,000	1,000	2,000	.49	1,587.60	-412.40	.62
11	3,300	100	1,100	1,000	2,100	.45	1,485.00	-615.00	.63
12	3,350	100	1,200	1,000	2,200	.41	1,373.50	-826.50	.66
13	3,395	100	1,300	1,000	2,300	.37	1,256.15	-1,043.85	.67
14	3,435	100	1,400	1,000	2,400	.33	1,133.55	-1,366.45	.70

whose situation was presented in Table 10—1A. Revenue opportunities are of a different kind, however. In that case we assumed the firm studied to be one among many producers of hose, each of the many producers turning out hose exactly like those produced by the others. In order to distinguish that case from the present one we shall now introduce a manufacturer of ties.

Suppose that the firm studied is either (a) the only producer of ties, (b) one of a very few producers of ties, or (c) one among many producers of ties each of whom turns out ties different from any of the others. In any of these three situations our tie producer would find that he could choose between selling more ties at a lower price per tie (a lower average revenue) or fewer ties at a higher price per tie (a higher average revenue). A statement of the price at which each output of ties could be sold is an *average revenue schedule*, a demand schedule for the product of the particular firm. In Table 10—3A this average revenue schedule is greater than the schedule assumed in Table 10—3B. Total revenue at each output is determined, as in the case of a purely competitive seller, by multiplying output by the price at which that output can be sold (that is by average revenue).

First let us examine the adjustments of this firm assuming the average revenue possibilities indicated in Table 10—3A. (1) A comparison between total revenue and total cost at each output shows that profits can easily be obtained and that the greatest possible profit is at an output of 2,908 ties, produced by the employment of 8 units of input. Profits at this output are \$3,389. (2) It is also evident that at many output levels total unit cost is less than the associated average revenue, and these are the same outputs at which total revenue exceeds total cost. (3) Such a firm does not, however, necessarily produce beyond the least-cost-combination, as would a purely competitive firm making a profit; monopolistic sellers making profits may produce at almost any level, greater or smaller or equal to the least-cost-combination position. (4) The price which this firm will charge is \$1.75; this is the price at which an output of 2,908 ties can be sold. It is thus the price at which profits are maximized. *The firm curtails production somewhat in order to get as high a price as this.*

Now let us look at the situation depicted in Table 10—3B. (1) A comparison between total revenues and total costs shows that at no output can this firm make a profit, though losses will be minimized at an output of 2,604 ties (that is 2,604 units of output), associated with an input of 7. (2) It is also evident that at no output levels is total unit cost less than average revenue; again we may conclude that there is no possibility of making a profit. (3) At several levels, however, total revenue is in excess of variable cost (and these are the same levels at which average revenue is greater than variable unit cost); it will therefore pay to continue to operate in the short run. (If average revenue were much less than this the firm would shut down.) (4) Such a firm will operate short of the least-cost-combination output. Proof of this fact will be taken up in a later chapter. (5) The price which the firm will charge is \$.61; this

is the price at which an output of 2,604 ties can be sold; it is the price at which losses are minimized. Again the *firm curtails production somewhat in order to get as high a price as this*. (6) The advantage in operating over shutting down is \$888.44; this is the excess of total revenue over variable cost at the least unprofitable production level. If the firm shut down it would get back nothing at all on its fixed costs of \$1,000; by operating it can cover the variable costs that will be incurred and in addition get back \$888.44 toward the meeting of fixed costs. The remaining sum of \$111.56, which is not covered, is a far smaller loss than the \$1,000 of fixed costs which would be lost if the firm shut down completely.

If average expenditure per unit of input were lower (or higher) than \$100, the cost data for the firm represented in these tables would be lower (or higher) and its profit position would therefore be more (or less) favorable. Its production and price adjustment would also be at a different level from that which we found in Tables 10—3A and 10—3B.

Graphic representation of profits and losses

The graphic representation of profits and losses of a monopolistic seller is similar to the graphic representation for a purely competitive seller.

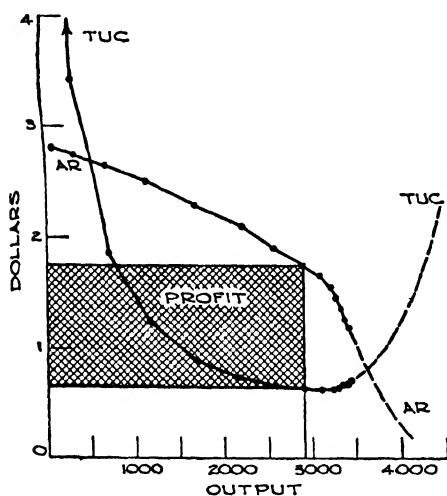


Fig. 10—2A. Monopolist making a profit.

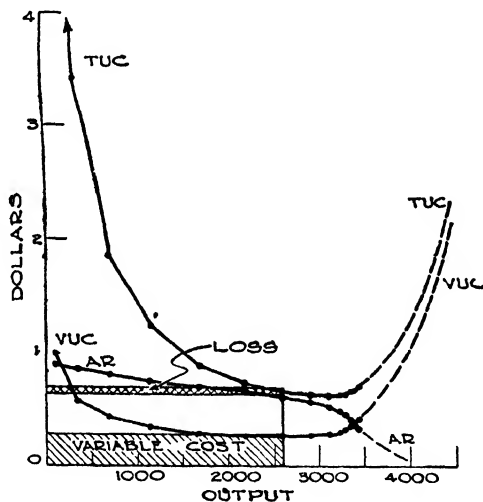


Fig. 10—2B. Monopolist operating at a loss.

Figure 10—2A is based on Table 10—3A, and Figure 10—2B is based on Table 10—3B. The average revenue line instead of being horizontal slopes down to the right, since the larger the output of the firm the lower will be the average revenue it can obtain by selling that output; that is, the lower will be the price it can get for its ties. In each case the most profitable output is again found by looking at the tables; and a vertical line is drawn at that output level. This vertical line intersects the line drawn to represent average revenue at just the price at which this output can be sold. A horizontal line drawn from this point on the average

revenue schedule will then mark off a rectangle which represents total revenue. Total cost and variable cost may be represented by rectangles marked off just as for a purely competitive firm. As in that case, a comparison between the rectangles for total revenue and those for total cost indicates profits or loss. And if a firm is operating at a loss the difference between total revenue and variable cost will also, as with a purely competitive firm, indicate the extent of the advantage in operating over shutting down.

The Policies of Monopsonistic Firms

A hypothetical monopsonistic buyer selling his product in a purely competitive market

The hose and the tie manufacturers of the previous illustrations were assumed to be buying labor services, raw materials, and other productive agents (or their services) in purely competitive markets. Let us now suppose that we are looking at the cost and revenue possibilities of a manufacturer of relatively cheap alarm clocks, that there are skilled workers employed by this manufacturer, and that he is the only entrepreneur (or one among a few entrepreneurs) in a certain area who is bidding for such skilled labor. In order to simplify our analysis we shall also make the artificial assumption that such labor is the *only* variable agent, so that a "unit of input" is identical with a unit of such labor services. We are further assuming that the manufacturer of clocks is selling his product in national markets in competition with a large number of other sellers offering clocks exactly like his; that is, we are assuming he sells clocks in a purely competitive market.

The monopsonistic buyer of skilled labor to produce alarm clocks may find that he can push down wages considerably provided that he is willing to get along with a small number of workers; but if he wants to attract men from other areas and other occupations into his firm he will have to raise his wages. He can get a large number of workers at a high wage, or a small number at a low wage. A schedule which states the wage he must pay per worker to attract each contemplated number of workmen is then his *average expenditure schedule* for labor.

Table 10—4 presents data for such a monopsonistic firm selling under purely competitive conditions. The average expenditure schedule tells us that this enterprise can get one worker for \$75 a month; it can get two if it pays an \$85 wage, three if it pays \$95, four if it pays \$105, and so on. Variable costs are equal to the expenditure per unit of input times the number of units of input. Thus two workers at \$85 each make variable costs two times \$85, or \$170; three workers at \$95 each make variable costs three times \$95, or \$285.

Just as in all the previous illustrations, this firm will determine its most profitable (or least unprofitable) production level by comparing total cost with total revenue at each level. A firm with cost and revenue possibilities like those shown in Table 10—4 would maximize profits by em-

TABLE 10—4

COST AND REVENUE DATA FOR A HYPOTHETICAL CLOCK MANUFACTURER:
BUYING MONOPSONISTICALLY; SELLING PURELY COMPETITIVELY

(1) <i>Input</i>	(2) <i>Output</i>	(3) <i>Average exp.</i>	(4) <i>Variable cost</i>	(5) <i>Fixed cost</i>	(6) <i>Total cost</i>	(7) <i>Average revenue</i>	(8) <i>Total revenue</i>	(9) <i>Profit</i>	(10) <i>Total unit cost</i>	(11) <i>Variable unit cost</i>
1	100	\$ 75	\$ 75	\$1,000	\$1,075	\$1.80	\$ 180	\$ -895	\$10.75	\$.75
2	350	85	170	1,000	1,170	1.80	630	-540	3.34	.49
3	702	95	285	1,000	1,285	1.80	1,264	-21	1.83	.41
4	1,152	105	420	1,000	1,420	1.80	2,074	654	1.23	.36
5	1,700	115	575	1,000	1,575	1.80	3,060	1,485	.93	.34
6	2,190	125	750	1,000	1,750	1.80	3,942	2,192	.80	.34
7	2,604	135	945	1,000	1,945	1.80	4,687	2,742	.75	.36
8	2,908	145	1,160	1,000	2,160	1.80	5,234	3,074	.74	.40
9	3,114	155	1,395	1,000	2,395	1.80	5,605	3,210	.77	.45
10	3,240	165	1,650	1,000	2,650	1.80	5,832	3,182	.82	.51
11	3,300	175	1,925	1,000	2,925	1.80	5,940	3,015	.89	.58
12	3,350	185	2,220	1,000	3,220	1.80	6,030	2,810	.96	.66
13	3,395	195	2,535	1,000	3,535	1.80	6,111	2,576	1.04	.75
14	3,435	205	2,870	1,000	3,870	1.80	6,183	2,313	1.13	.84

playing 9 skilled clock makers, producing 3,114 clocks. At this level profits are \$3,210.

If the average expenditures necessary to get clock makers were greater than indicated in Table 10—4 perhaps profits could not be made, but so long as there was any adjustment in which variable costs would be less than total revenue the clock manufacturer would continue in business in the short run.

In finding the most profitable (or least unprofitable) production level for this clock-manufacturing firm, *we have also found what wage will be paid.* With costs and revenues as given in Table 10—4, the wage associated with the most profitable adjustment is \$155 per month. *Inputs of workers are restricted in order to hold wages down,* and there is therefore an associated curtailing of output.

A monopsonistic buyer selling monopolistically

The firm buying the services of the skilled clock makers might have been in a monopolistic position in selling its clocks. In order to distinguish the example of the clock manufacturer as discussed in the last section from an example of monopolistic selling along with monopsonistic buying, we shall introduce another hypothetical firm. We shall assume a manufacturer of hand-carved boxes to be monopsonistic in buying productive agents, and that his costs and his buying position are exactly like those of the entrepreneur of the clock factory. But this box manufacturer is for some reason monopolistic in the sale of his boxes. His revenue opportunities would then resemble the position of the sellers of ties. Combining the clock manufacturer's monopsonistic situation with the tie manufacturer's monopolistic situation, we may use their data to study the box-making concern. These data are brought together in Table 10—5. The basic comparison between total revenues and total costs, and between total revenues and variable costs, may then be made as before. The most profitable production level in this case is found to be at 7 units of input giving an output of 2,604 boxes. Profits at that level are \$3,002.60. *Having found the most profitable (or least unprofitable) production level, we have also found (1) what wages will be paid, and (2) what price will be charged per box.* In this case the wage (average expenditure) will be \$135; the price per box will be \$1.90. *This entrepreneur curtails production both in order to raise the price of what he sells and to lower the price (wages, and so on) of what he buys.*

Monopolistic Sellers Facing Monopsonistic Buyers

In all of the preceding sections of this chapter there has been a tacit assumption that the firm concerned stood opposite customers who were purely competitive in buying from it and sellers who were purely competitive in selling to it. Pure competition is in fact characteristic among buyers who are ultimate consumers, and was an implied assumption in the discussion of consumer demand in Chapter 3. It is also approximated among laborers selling their services in labor markets where there is no

TABLE 10—5

COST AND REVENUE DATA FOR A HYPOTHETICAL BOX MANUFACTURER:
BUYING MONOPSONISTICALLY; SELLING MONOPOLISTICALLY

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
<i>Input</i>	<i>Output</i>	<i>Average exp.</i>	<i>Variable cost</i>	<i>Fixed cost</i>	<i>Total cost</i>	<i>Average revenue</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Total unit cost</i>	<i>Variable unit cost</i>
1	100	\$ 75	\$ 75	\$1,000	\$1,075	\$2.80	\$ 280.00	\$-795.00	\$10.75	\$.75
2	350	85	170	1,000	1,170	2.75	962.50	-207.50	3.34	.49
3	702	95	285	1,000	1,285	2.65	1,860.30	575.30	1.83	.41
4	1,152	105	420	1,000	1,420	2.50	2,880.00	1,460.00	1.23	.36
5	1,700	115	575	1,000	1,575	2.30	3,910.00	2,335.00	.93	.34
6	2,190	125	750	1,000	1,750	2.10	4,599.00	2,849.00	.80	.34
7	2,604	135	945	1,000	1,945	1.90	4,947.60	3,002.60	.75	.36
8	2,908	145	1,160	1,000	2,160	1.75	5,089.00	2,929.00	.74	.40
9	3,114	155	1,395	1,000	2,395	1.65	5,138.10	2,743.10	.77	.45
10	3,240	165	1,650	1,000	2,650	1.55	5,022.20	2,372.00	.82	.51
11	3,300	175	1,925	1,000	2,925	1.45	4,785.00	1,860.00	.89	.58
12	3,350	185	2,220	1,000	3,220	1.35	4,522.50	1,302.50	.96	.66
13	3,395	195	2,535	1,000	3,535	1.25	4,243.75	708.75	1.04	.75
14	3,435	205	2,870	1,000	3,870	1.15	3,950.25	80.25	1.13	.84

trade union, and is typical of conditions among sellers of many of the raw materials produced in agricultural enterprise. It is approximated very closely by manufacturers of a wide range of products from shoes to hats, though strictly speaking these cases might be regarded as monopolistic competition.

There are, however, many important cases in which monopolistic sellers are offering their products (or services) in markets in which buyers of these products (or services) are monopsonistic. There will be a minimum price at which a monopolistic seller will be willing to sell any given quantity if he can do no better than this; there will on the other hand be a maximum price at which a monopsonistic buyer will be willing to purchase any given quantity. Higgling concerning price and quantity will result in a bargained price which is somewhere between this seller's minimum and the buyer's maximum for the quantity exchanged. While the solution is not determinate, and depends on the skill and strength of the parties to the bargain, a more detailed analysis of such adjustments can be given when we have further sharpened our analytical tools.

Price Discrimination

Occasionally a firm is in a position to charge different prices to different customers, that is to "discriminate" between different customers. For example, railroads in the past granted rebates to large shippers, the most notorious example being that of the Standard Oil Company; ordinary shippers under these conditions paid higher freight rates than did Mr. Rockefeller. The railroads also charged higher rates on freight shipments between points that did not offer an alternative of water transportation than the rates charged on rail hauls from one port to another.

There is also occasionally price discrimination in the amounts paid for the same goods (or services) bought from different sellers. Packing houses commonly buy cream and make butter for sale. In order to break some of the dairy cooperatives the buyers from these packing houses have offered high prices to dairy farmers in areas where cooperative creameries were competing to buy cream, while these same packing houses were paying lower prices for cream in adjoining areas.

The tacit assumption of the preceding sections of this chapter was that a firm sold at the same price to all customers and bought at the same price from all those from whom it made purchases. That assumption will be continued in most of the following chapters. Some important problems associated with discriminatory pricing practices will, however, be considered in Chapter 26.

CHAPTER 11

Introduction to Marginal Adjustments

WHEN John is trying to make up his mind as to how to spend his income he does not ask himself, "Shall I buy 30 pairs of socks and no shoes, or 5 pairs of shoes and no socks?" Such a question seems absurd. He is more likely to ask, "Would I be better satisfied with 3 pairs of shoes and 15 pairs of socks or with 2 pairs of shoes and 20 pairs of socks?" Perhaps he has usually purchased in such a way as to keep on hand 3 pairs of shoes and 15 pairs of socks, and he is contemplating whether it would add more to his satisfactions to buy another pair of shoes and refrain from replacing some socks or to buy more socks and refrain from replacing one of his pairs of shoes. These are *marginal* comparisons, they are comparisons between a little more of this and a little less of that, or a little less of this and a little more of that.

An entrepreneur making long-run plans and seeking to arrive at the most efficient combination of productive agents will not ask, "Shall we hire 50 laborers and do without plant and machinery, or shall we build a big plant and do without laborers?" This again would be an absurd question. Instead the entrepreneur might ask himself whether it would be better to incur a cost of \$1,000,000 for equipment, which will amount to about \$1,000 a month depreciation cost, and spend \$2,000 a month on labor, or to incur a cost of \$1,100,000 for equipment, which will amount to about \$1,100 a month depreciation cost, and spend \$1,900 a month on labor. These again are *marginal* comparisons: they are comparisons between a little more of this and a little less of that, or a little less of this and a little more of that.

Under ordinary circumstances*the entrepreneur of a going concern will not question whether to produce no shoes or 1,000,000 pairs a month (unless this is a large concern just contemplating shutting down because business is so bad). Rather he will question whether it will be better to plan on production of 10,000 shoes next month or 11,000; whether to hire 14 shoe workers or 15 shoe workers. He will compare the difference in labor cost with the difference in total revenues resulting from the alternative courses of action. This too involves a *marginal* decision, it probably involves a comparison between the advantage of higher revenues and the disadvantage of greater costs.

When people say that we have "too much wheat and too little housing"

they are not suggesting that we do away with wheat production entirely and concentrate our efforts on housing. Rather they are suggesting that our resources would be better used if *some* of the labor and other resources going into the production of wheat could be diverted to other uses, especially house construction. They are suggesting that a little more housing and a little less wheat would be a good thing. They are making *marginal* comparisons in evaluating the way in which resources are allocated in our economy.

These rough examples are sufficient to indicate the great importance of marginal comparisons in studying and evaluating economic adjustments. Such comparisons will be made continuously in subsequent discussions in this book. In this chapter certain aspects of the "marginal concept" will be explored; we shall examine short-run costs and revenues of a firm from this point of view. In the following chapters we shall find that through this approach we shall arrive at the same answers concerning the policies of firms that we reached in the last chapter. But we shall have gained added insight into the significance of the differences in the adjustments of purely competitive firms and of firms in monopolistic or monopsonistic positions.

Defining Some Marginal Concepts

When an entrepreneur adds a unit of input (let us say a "package" of one laborer and associated raw materials) several results will follow: (1) outputs will be increased; (2) costs will be increased; and (3) total revenue of the firm will probably be different from that obtained in selling a smaller amount of product. Each of these changes involves a *marginal* change, an addition (or subtraction). The three marginal concepts directly associated with such a change are:

1. *Marginal product*; the difference between two successive total products (e.g., the difference between total products resulting from ten and from eleven units of input). This concept has already been described in Chapter 6.

2. *Marginal cost*; the difference between two successive total costs (e.g., the difference between total costs when ten and when eleven units of input are employed).

3. *Marginal revenue*; the difference between two successive total revenue possibilities (e.g., the difference between total revenues from the sale of the product of ten and the product of eleven units of input).

An extremely simplified example will bring out the nature of these relations. Suppose a certain firm to be producing cardboard picture puzzles, and that the results of different inputs are as shown in Table 11—1. In this case adding the fifth unit of input raises output from 1,000 to 1,200 puzzles; marginal product is 200. The fifth unit of input raises total cost from \$400 to \$525; marginal cost is \$125. And this fifth unit of input raises the total revenue obtainable by selling the puzzles from \$1,000 to \$1,140; marginal revenue is \$140. Adding the sixth unit of input raises

TABLE 11—1

MARGINAL DATA FOR A HYPOTHETICAL PUZZLE-MANUFACTURING FIRM

<i>Input</i>	<i>Total product (output)</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Marginal product</i>	<i>Marginal cost</i>	<i>Marginal revenue</i>
4	1,000	\$400	\$1,000	?	\$?	\$?
5	1,200	525	1,140	200	125	+140
6	1,350	660	1,215	150	135	+75
7	1,450	805	1,160	100	145	-55

output from 1,200 to 1,350; marginal product is 150. It raises total cost from \$525 to \$660; marginal cost is \$135. It raises total revenue from \$1,140 to \$1,215; marginal revenue is \$75. Adding the seventh unit of input raises output from 1,350 to 1,450; marginal product is 100. It raises total cost from \$660 to \$805; marginal cost is \$145. It lowers total revenue from \$1,215 to \$1,160 (because the greater output can be sold only for a sharply reduced price); marginal revenue is a negative quantity, that is, -\$55.

Marginal Costs and Average Expenditures of Purely Competitive and of Monopsonistic Buyers

Purely competitive buying

When the entrepreneur of a firm buying purely competitively adds a unit to input, he increases costs by exactly the payment he makes to the new agents; that is, *in purely competitive buying marginal cost at each input level equals average expenditure.* This is true because a firm which is purely competitive in the buying of productive services will pay the same amount per unit of input whatever the amount purchased; that is, "average expenditure" will be the same at all inputs.

Such a situation is illustrated in Table 11—2, which is based on the

TABLE 11—2

COSTS OF A HYPOTHETICAL TIE MANUFACTURER BUYING PURELY COMPETITIVELY

(1) <i>Input</i>	(2) <i>Average expenditure</i>	(3) <i>Variable cost</i>	(4) <i>Fixed cost</i>	(5) <i>Total cost</i>	(6) <i>Marginal cost</i>
1	100	\$ 100	\$1,000	\$1,100	\$...
2	100	200	1,000	1,200	100
3	100	300	1,000	1,300	100
4	100	400	1,000	1,400	100
5	100	500	1,000	1,500	100
6	100	600	1,000	1,600	100
7	100	700	1,000	1,700	100
8	100	800	1,000	1,800	100
9	100	900	1,000	1,900	100
10	100	1,000	1,000	2,000	100
11	100	1,100	1,000	2,100	100
12	100	1,200	1,000	2,200	100
13	100	1,300	1,000	2,300	100
14	100	1,400	1,000	2,400	100

same data as were given for the hypothetical tie-manufacturing enterprise (and for the hose manufacturer) discussed in Chapter 10. When this tie manufacturer adds a unit of input (a "package of variable agents") he pays to the new agents \$100. This \$100 is his "average expenditure" at all input levels. When he adds a unit to input he adds \$100 to variable costs, and total costs are therefore also increased by \$100; that is marginal cost in each case is also \$100. When this tie manufacturer considers what the effects on cost of employing another unit of input would be, he finds that the amount he will pay to the new agents is exactly what he will add to his production costs; average expenditure and marginal cost are equal.

Monopsonistic buying

When a monopsonist adds a unit to input he increases his costs by more than the payments to the new agents; that is, *in monopsonistic buying marginal cost at each input level exceeds average expenditure*. Here we may draw on the case of the clock manufacturer (or the box producer) discussed in Chapter 10. Table 11—3 is based on the data for that enter-

TABLE 11—3

COSTS OF A HYPOTHETICAL CLOCK MANUFACTURER BUYING MONOPSONISTICALLY

(1) Input	(2) Average expenditure	(3) Variable cost	(4) Fixed cost	(5) Total cost	(6) Marginal cost
1	\$ 75	\$ 75	\$1,000	\$1,075	\$...
2	85	170	1,000	1,170	95
3	95	285	1,000	1,285	115
4	105	420	1,000	1,420	135
5	115	575	1,000	1,575	155
6	125	750	1,000	1,750	175
7	135	945	1,000	1,945	195
8	145	1,160	1,000	2,160	215
9	155	1,395	1,000	2,395	235
10	165	1,650	1,000	2,650	255
11	175	1,925	1,000	2,925	275
12	185	2,220	1,000	3,220	295
13	195	2,535	1,000	3,535	315
14	205	2,870	1,000	3,870	335

prise. In order to get 6 skilled clock makers the entrepreneur of that firm had to pay them \$125 a month wage; he could still retain 5 workers if he reduced the wage to \$115 a month. But the difference between the cost of employing 5 and the cost of employing 6 workers is more than the \$125 paid the sixth worker; that is, the resulting marginal cost is greater than the average expenditure when 6 workers are employed.

The reason for this is in fact simple: *each increase in input of skilled clock makers was obtainable only by paying for all units (workers) a price (wage) higher than before*. The clock manufacturer could not very well bid for additional workers by offering them higher pay without at the same time raising the pay of all the workers of the same grade and effi-

ciency employed in his shop. His wage bill for 6 workers exceeds that for 5 workers not only because \$125 is spent on the sixth worker, but because he must pay an additional \$10 a month to the first 5 workers. This means an increase in pay roll of \$125 *plus* 5 times \$10, or a total increase of \$175. The marginal cost at 6 units of input is \$175; the average expenditure is only \$125.

Graphic representation of the relation between marginal cost and average expenditure

The relation between marginal cost and average expenditure in purely competitive and in monopsonistic situations may be represented graphically by plotting numbers of units of *inputs* along the horizontal axis.

Figure 11—1A is based on Table 11—2; both average expenditure (*AE*)

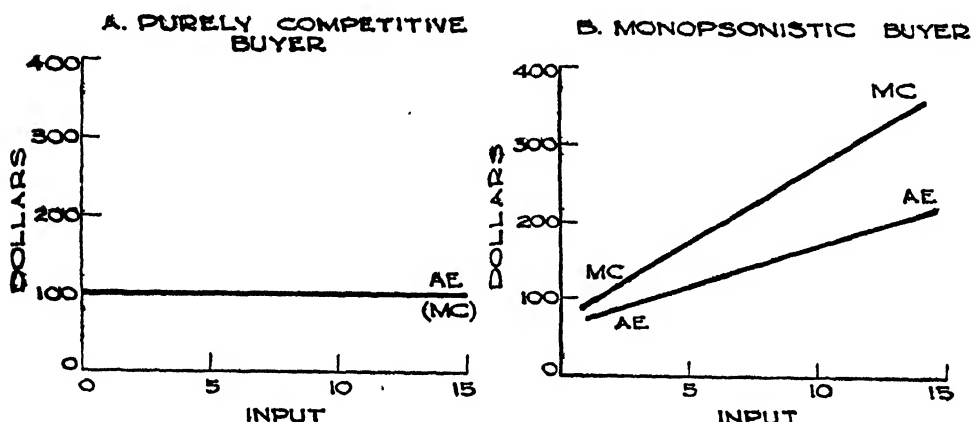


Fig. 11—1. Marginal cost and average expenditure.

and marginal cost (*MC*) are \$100 at all inputs, and are represented by a horizontal line at the level of \$100. This is the way marginal costs and average expenditures would look to a purely competitive buyer. Figure 11—1B is based on Table 11—3; average expenditure (*AE*) varies with the number of units of input, indicating that the firm is buying under monopsonistic conditions. It can push down payments per unit of input if it is willing to curtail employment; it can expand only by paying more for the services of these agents. As a result marginal cost (*MC*) is consistently higher than average expenditure.

Average expenditure as a measure of the "worth" of a unit of input in other uses

Average expenditure may be regarded as a rough measure of the "worth" of a unit of input in *other* uses. If in order to attract 10 workers it is necessary to pay wages of \$150 a month, this is a rough indication that in other firms and in other industries there are opportunities for employment of one more worker at about as much as this. Otherwise the wage which would have to be paid to attract the tenth worker here would

fall. It is even more definitely an indication that none of the 10 workers have other opportunities that would bring them more than \$150 a month: if they did they would probably take advantage of these more lucrative alternatives. Their worth in alternative employments reflects indirectly the demands of consumers for other products and hence for the services of workers in the production of these other goods. Despite its crudeness, this wage measure is a useful tool in evaluating the policies of firms and the allocation of workers among them. And what is true of the price of labor is probably even more true of the prices of other productive agents (or their services). *The average expenditure on a unit of input in a particular firm is thus an index of the worth of one unit of input if the agents involved were diverted to other enterprises.*

Marginal Revenue and the Sale Value of the Marginal Products of Purely Competitive and of Monopolistic Sellers

Marginal revenue and the elasticity of an average revenue schedule

When an entrepreneur decides to expand the level of production, that decision involves not only additions to costs, but additions to output as well. Before he undertakes such an expansion an entrepreneur will consider the expected effect on his total revenue of selling the larger instead of the smaller output.

Sometimes total revenue will be larger as the sales of the firm are expanded; sometimes it will be smaller. This will depend on whether the firm can expand its sales considerably with only a relatively small reduction in price or whether, on the other hand, expansion of sales can be made only by marked reductions in the price per unit sold. Whenever consumers are so responsive to a reduction in price that they will spend more money at a lower price, the firm will find its total revenues increased by expanding output and reducing price. Whenever consumers are unresponsive to a reduction in price, so that their increased consumption is very small and they spend less on the good sold by this firm at a lower than at a higher price, the firm will find its total revenues diminished by expanding output and reducing price. In other words, when the average revenue schedule is "elastic" total revenue will be greater at the lower price and expanded output; when the average revenue schedule is "inelastic" total revenue will be less at the lower price and expanded output. *Marginal revenue is positive when the average revenue schedule is elastic; marginal revenue is negative when the average revenue schedule is inelastic.*

Since in the extreme case of the purely competitive seller output of the firm can be expanded without exerting enough influence in the total picture to require any reduction in price at all, in this case total revenue is of course always larger the larger the quantity of product sold. The average revenue schedule of a purely competitive seller is *infinitely elastic*, and marginal revenue is positive for all expansions of production.

These relations are easily illustrated by the use of graphs, as in the analysis of demand in Chapter 3. In fact an average revenue schedule is a special kind of demand schedule; it is a schedule of the demand for the product of a single seller. In Figures 11—2A and 11—2B outputs of the

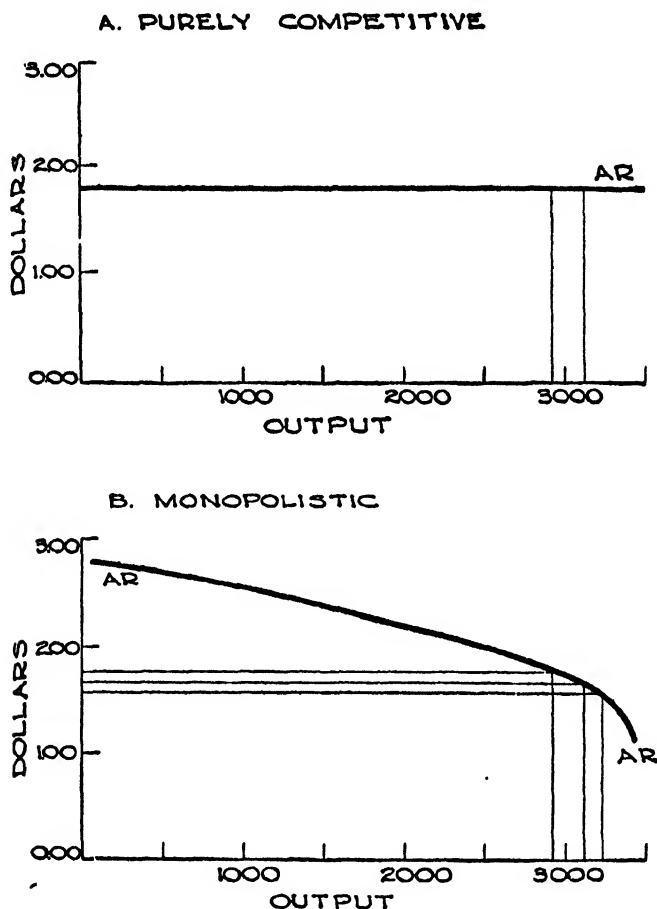


Fig. 11—2. Average revenue of a hypothetical seller.

firm are marked off along the horizontal axis, and the prices at which these outputs could be sold are marked off vertically. Figure 11—2A represents the average revenue schedule of a purely competitive seller; whatever the quantity sold he gets the same price of \$1.80. In fact, this might be a graphic representation of the position of the hose manufacturer (or the clock manufacturer) of the illustrations of Chapter 10. The total revenue obtained from the sale of 2,908 pairs of hose is represented by a rectangle the height of which is the price of \$1.80 and the length of which is the sales of 2,908 pairs of hose (units of output). The total revenue from any larger output, say 3,114 units, is similarly represented. Total

revenue at the larger outputs is obviously always greater than at the smaller outputs.

Figure 11—2B represents the average revenue schedule of a monopolistic seller; larger quantities bring lower prices, while small quantities can be sold at higher prices. This might be a graphic representation of the position of the tie manufacturer (or of the carved-box manufacturer) of the illustrations of Chapter 10. The average revenue schedule shows that an output of 2,908 ties can be sold for a price of \$1.75 each. The height of a vertical line drawn from this point on the average revenue schedule to the base line indicates this price of \$1.75; the length of a horizontal line back to the vertical axis indicates the output; the rectangle marked off by these lines indicates total revenue. A similar rectangle showing total revenue at an output of 3,114 ties is larger than this; but the rectangle showing total revenue at an output of 3,240 ties is smaller than that showing total revenue at an output of 3,114.

Marginal revenue and the sale value of the marginal product

When an entrepreneur adds a unit to input he expects to increase his total product as a result; this addition to total product is what we have called marginal product. Thus, in a dairy if another cow (plus feed and so on) adds 1,000 quarts of milk a year the marginal product is then 1,000 quarts of milk. The number of dollars taken in from customers in exchange for these 1,000 quarts of milk is the "sale value of the marginal product." This is a cumbersome phrase, but it says exactly what it means. *The sale value of the marginal product is the number of dollars for which the marginal product will sell.* If milk is selling at 12 cents a quart, the sale value of a marginal product of 1,000 quarts of milk would be 1,000 times \$.12, or \$120. What is the relation between this sale value of the marginal product and the marginal revenue resulting from expansion of output by the employment of an additional unit of input?

Under conditions of pure competition in selling, marginal revenue and the associated sale value of marginal product will always be equal. If a firm is purely competitive in selling, the price it will get will be the same whatever its output. This situation is illustrated in Table 11—4, which is based on the same figures as were used in the illustrations of the hose (and the clock) manufacturer of Chapter 10. Whatever the output of hose, this producer gets \$1.80 a pair. For example, when the tenth unit of input is added, output is increased by 126 units. The additional 126 pairs of hose still sell for \$1.80 a pair; the sale value of the marginal product resulting from employment of the tenth unit of input is then 126 times \$1.80, or \$226.80. But this is exactly equal to the difference between the total revenue obtainable from selling the 3,240 pairs of hose produced by 10 units of input and the 3,114 pairs of hose produced by 9 units of input: that is, it is equal to the marginal revenue resulting from the expansion in production.

The case of a monopolistic seller is different. *Under monopolistic conditions the sale value of the marginal product will always be greater than*

TABLE 11—4

REVENUE SCHEDULES OF A HYPOTHETICAL PURELY COMPETITIVE SELLER OF HOSE

<i>Input</i>	<i>Output (sales)</i>	<i>Marginal product</i>	<i>Average revenue</i>	<i>Total revenue</i>	<i>Marginal revenue</i>	<i>Sale value of marginal product</i>
1	100	100	\$1.80	\$ 180.00	\$180.00	\$180.00
2	350	250	1.80	630.00	450.00	450.00
3	702	352	1.80	1263.60	633.60	634.60
4	1152	450	1.80	2073.60	810.00	810.00
5	1700	548	1.80	3060.00	986.40	986.40
6	2190	490	1.80	3942.00	882.00	882.00
7	2604	414	1.80	4687.20	745.20	745.20
8	2908	304	1.80	5234.40	547.20	547.20
9	3114	206	1.80	5605.20	370.80	370.80
10	3240	126	1.80	5832.00	226.80	227.80
11	3300	60	1.80	5940.00	108.00	108.00
12	3350	50	1.80	6030.00	90.00	90.00
13	3395	45	1.80	6111.00	81.00	81.00
14	3435	40	1.80	6183.00	72.00	72.00

the associated marginal revenue. Here we shall use as an example the tie manufacturer who has already served us on a number of occasions. This tie manufacturer's situation is reproduced in Table 11—5, the data being

TABLE 11—5

REVENUE SCHEDULES OF A HYPOTHETICAL MONOPOLISTIC SELLER OF TIES

<i>Input</i>	<i>Output (sales)</i>	<i>Marginal product</i>	<i>Average revenue</i>	<i>Total revenue</i>	<i>Marginal revenue</i>	<i>Sale value of marginal product</i>
1	100	100	\$2.80	\$ 280.00	\$ 280.00	\$ 280.00
2	350	250	2.75	962.50	682.50	687.50
3	702	352	2.65	1860.30	897.80	932.80
4	1152	450	2.50	2880.00	1019.70	1125.00
5	1700	548	2.30	3910.00	1030.00	1260.40
6	2190	490	2.10	4599.00	689.00	1029.00
7	2604	414	1.90	4947.60	348.60	786.60
8	2908	304	1.75	5089.00	141.40	532.00
9	3114	206	1.65	5138.10	49.10	339.90
10	3240	126	1.55	5022.00	-116.10	195.30
11	3300	60	1.45	4785.00	-237.00	87.00
12	3350	50	1.35	4522.50	-262.50	67.50
13	3395	45	1.25	4243.75	-278.75	56.25
14	3435	40	1.15	3950.25	-293.50	46.00

based on those of the illustration in Chapter 10. The larger the output of ties, the lower the price that can be obtained. Suppose this entrepreneur decides to try to sell 2,604 ties instead of 2,190 ties (produced by increasing his inputs from 6 to 7). When he passes the additional 414 ties over to customers he receives from these customers \$1.90 per tie: the sale value of the marginal product (of 414 ties) is 414 times \$1.90 or \$786.60.

But his total revenue is not increased by this full \$786.60. When he tries to sell 2,604 ties instead of 2,190, he is forced to reduce the price on *all* his ties and not merely on the marginal product of 414 units. The reduction of \$.20 per tie on the initial 2,190 ties must be taken into account, and this amounts to \$438. The net addition to his total revenue is therefore only \$786.60 minus \$438 or \$348.60. This checks with the figures in the marginal revenue column, obtained by making a direct comparison between total revenue from the sale of 2,190 ties and from the sale of 2,604 ties. *The sale value of the marginal product resulting from expanding production in a monopolistic firm is always greater than the associated marginal*

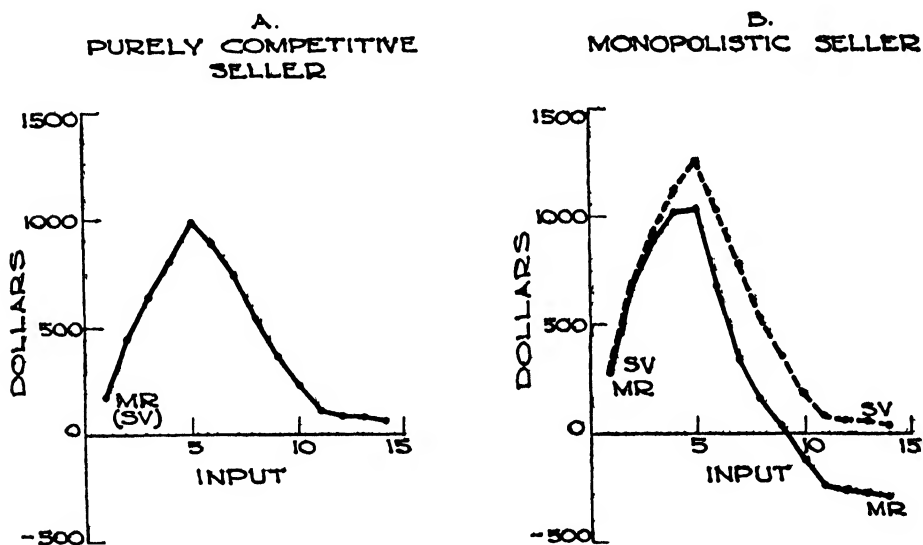


Fig. 11—3. Marginal revenue and sale value of a marginal product.

revenue because of the fact that it is necessary to reduce the price on all the units sold, and not merely on those making up the marginal product.

These relations between marginal revenues and the sale values of marginal products may be illustrated graphically. Such illustration adds nothing fundamentally new to the analysis, as is true of all graphic illustration, but it may prove useful by presenting a visual picture of the relationship. In Figures 11—3A and 11—3B inputs are marked off along the base line and marginal revenues and the sale values of marginal products are then plotted for each input level. Figure 11—3A is based on the data of Table 11—4, representing a purely competitive seller. Marginal revenue and the associated (and equal) sale values of the marginal products rise and then fall as marginal product itself rises and then falls. Figure 11—3B, based on the data of Table 11—5, represents a monopolistic seller. Again marginal revenues and sale values of the marginal products rise and then fall, reflecting changing marginal products; but in this case there are two separate curves. The curve represent-

ing the sale values of the marginal products is consistently above that representing the associated marginal revenues.

**"Sale value of the marginal product" as a measure of the
"worth" of the marginal product to consumers**

The sale value of the marginal product resulting from any given expansion in production in a firm provides a rough measure of the worth to consumers of that marginal product. Thus if the tie manufacturer of Table 11—4 were to expand production up to 3,240 ties, the additional 126 ties resulting from the employment of the tenth unit of input would sell for \$1.55 apiece, totaling \$195.30. Consumers are clearly willing to pay at least \$1.55 apiece to get these additional 126 ties. They would not have been willing to pay as much as \$1.65 apiece for them. The sale value of the marginal product (of 126 ties in this case) is a measure which shows us that the aggregate of consumers considered the added output of 126 ties to be worth at least as much as \$195.30 worth of other things. Otherwise they would spend their money on the other things.

It is for this reason that we have laid emphasis on the difference between marginal revenue and the sale value of the marginal product. The marginal revenue resulting from expanding output of ties from 3,114 to 3,240 units is actually negative; so far as this entrepreneur is concerned 3,240 ties are worth less than 3,114 since he cannot sell the larger quantity for as much as the smaller quantity. In fact he would rather destroy some ties than sell as many as 3,240 at the low price he could get for them. He is concerned primarily with what happens to his costs and revenues, not with consumers' welfare. Yet it would be absurd to say that consumers would be worse off if some philanthropist were to present them with 3,240 ties free than if he were to give them the smaller quantity of 3,114 ties! Here there is clearly a conflict between the interests of the tie producer and of the consumers of ties. Many conflicts are less obvious than this; but as we shall see, the profit motive as it works out in a monopolistic enterprise generally tends to lead to policies which are to some degree incompatible with consumer interests.

CHAPTER 12

Marginal Analysis of the Most Profitable Level of Production

WE ARE NOW ready to engage in a little magic, to show that by comparing marginal revenues with marginal costs we shall arrive at conclusions concerning the most profitable levels of production that are exactly the same as those reached by comparing total revenues with total costs in Chapter 10. But when we have traced through these relations we shall have something more than the conclusions reached in that chapter. We shall have a measure of the extent to which the behavior of firms in different types of market situations will tend to be consistent or inconsistent with that allocation of resources which would most nearly fulfill the preferences of consumers. In order to make our analysis directly comparable with that of Chapter 10 we shall repeat the hypothetical cases that we used there and that we have drawn upon in Chapter 11.

In any going concern the major decisions are essentially "marginal" decisions: they involve questions as to whether inputs should be increased or decreased. In making such decisions the crucial comparisons are between the changes which would ensue in costs on the one hand, in revenues on the other. Whenever an addition to input would add more to revenue than to cost, it would clearly improve the position of the firm; when it would add more to cost than to revenue it would put the firm in a less favorable profit position than before. To express this in technical terminology: *the maximum profit (or minimum loss) adjustment will be attained by adding units to input so long as marginal revenue continues to exceed marginal cost.* Whether the result is a profit or a loss, it will still be the best possible adjustment open to the firm. This generalization applies to firms in all types of buying and selling positions, despite the important differences between them.

Adjustments in Purely Competitive Firms

Marginal analysis of the most profitable level of production

The simplest type of situation which a firm may face is that of pure competition in the buying of productive services and in the selling of its product. We shall therefore turn once more to our now familiar friend,

the hypothetical purely competitive manufacturer of hose. In Table 12—1 we have reproduced figures for inputs, outputs, total costs, total revenues, average expenditures, average revenues, and profits, as they were entered in Table 10—1A. To this table we have added columns giving marginal costs, marginal revenues, marginal products, and the sale values of these marginal products. We shall ask the cooperation of the reader in pretending the columns giving total cost, total revenues, and profits are not there until we suggest that he look at them.

Let us suppose that this entrepreneur has started production experimentally **at an input of 9 units, resulting in an output of 3,114 pairs of hose**. He can get \$1.80 per pair of hose whatever the quantity of hose he may sell. A glance at the marginal revenue and the marginal cost that would result from expanding his production by adding another unit to input tells us immediately that he would improve his position by such an expansion. He would add \$227 to his revenues and only \$100 to his costs; no matter what his profit position at an input of 9 units and an output of 3,114 units may be, he would clearly gain by raising his level of production. In fact he is still not making the most of his opportunities if he stops at 10 units of input producing 3,240 pairs of hose, since by adding the eleventh unit to input he would add \$108 to his revenues and only \$100 to his costs. He would not, however, gain anything from expanding beyond the input of 11 units, producing 3,300 units of output. The addition of the twelfth unit to input would increase his revenues by only \$90 and his costs by \$100 so that he would not be in as favorable a position as before. He could maximize his profits (or minimize his losses) by producing at a level of 11 units of input, turning out a total production of 3,300 pairs of hose.

The reader is now given permission to look at the total cost, the total revenue, and the profit columns. What appears? The differences between total revenue and total cost at the various production levels show that the most profitable position is at 11 units of input and an output of 3,300 pairs of hose. Our answers check.

Evaluation of the level of production in a purely competitive firm

What is it that this marginal analysis has revealed that was not taken care of simply by comparing total costs with total revenues? It has shown us three significant things:

1. Marginal revenues equal the sale values of the marginal products; this purely competitive seller has therefore expanded output so long as the addition to his cost was less than the sale value of the marginal product; *he has expanded output so long as the addition to his cost was less than the worth to consumers of the increase in output*. In this respect he has acted consistently with consumers' interests even though his purpose was merely to maximize his own profits.

2. Marginal costs equal average expenditures; this purely competitive buyer has therefore expanded input so long as the addition to his revenues

TABLE 12—1
THE COSTS AND REVENUES OF A HYPOTHETICAL PURELY COMPETITIVE MANUFACTURER OF HOSE

<i>Input</i>	<i>Output</i>	<i>Average revenue</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Average expenditure</i>	<i>Marginal cost</i>	<i>Marginal revenue</i>	<i>Marginal product</i>	<i>Sale value of marginal product</i>
1	100	\$1.80	\$1,100	\$ 180.00	\$ -920.00	\$100	\$100	\$180.00	100	\$180.00
2	350	1.80	1,200	630.00	-570.00	100	100	450.00	250	450.00
3	702	1.80	1,300	1,263.60	-36.40	100	100	633.60	352	633.60
4	1,152	1.80	1,400	2,073.60	673.60	100	100	810.00	450	810.00
5	1,700	1.80	1,500	3,060.00	1,560.00	100	100	986.40	548	986.40
6	2,190	1.80	1,600	3,942.00	2,342.00	100	100	882.00	490	882.00
7	2,604	1.80	1,700	4,687.20	2,987.20	100	100	745.20	414	745.20
8	2,908	1.80	1,800	5,234.40	3,434.40	100	100	547.20	304	547.20
9	3,114	1.80	1,900	5,605.20	3,705.20	100	100	370.80	206	370.80
10	3,240	1.80	2,000	5,832.00	3,832.00	100	100	226.80	126	226.80
11	3,300	1.80	2,100	5,940.00	3,840.00	100	100	108.00	60	108.00
12	3,350	1.80	2,200	6,030.00	3,830.00	100	100	90.00	50	90.00
13	3,395	1.80	2,300	6,111.00	3,811.00	100	100	81.00	45	81.00
14	3,435	1.80	2,400	6,183.00	3,783.00	100	100	72.00	40	72.00

was greater than the sum he would pay to the agents included in the additional unit of input; *he has continued to hire additional workers and to buy additional raw silk as long as the wage he pays for the labor plus the price he pays for the silk is less than a unit of input would add to his revenues.* In this respect he has acted consistently with the interests of the sellers of labor and of raw silk, even though his purpose was merely to maximize his own profits; and he has acted consistently with consumer preferences in bidding these additional agents into his firm and away from less important uses.

3. Since marginal costs equal average expenditures, and marginal revenues equal the sale values of the marginal products, this purely competitive entrepreneur has continued to expand his level of production so long as average expenditure was less than the sale value of the marginal product. *He was therefore adding workers and other agents in his firm so long as the worth to consumers of their additions to his product exceeded what entrepreneurs of other firms would bid for their services.* Insofar as entrepreneurs of other firms are also purely competitive, their bids reflect consumer demands for their products. He is therefore carrying production in his firm just to the point which is most consistent with consumer preferences concerning allocation of resources to this as compared with other uses. So long as consumers wanted a little more of labor services used in producing hose even at a little sacrifice of these labor services used in producing wheat, he has complied with their wishes; he has not gone beyond that point. *Without being in the least a philanthropist, this purely competitive entrepreneur seeking to maximize profits has fulfilled very effectively a service in helping to allocate resources consistently with consumer preferences.*

Adjustments in Firms Which Are Not Purely Competitive in Both Buying and Selling

A hypothetical firm that buys purely competitively and sells monopolistically

The marginal analysis of the most profitable level of production in a hypothetical purely competitive hose manufacturing enterprise provides us with a background for a similar analysis of the hypothetical tie producer whose situation we developed as an illustration in Chapter 10. This tie manufacturer was assumed to be buying all productive services in purely competitive markets, but selling his ties in a market in which (for any of a number of possible reasons) he has some degree of individual control over price. He can choose to sell many ties at a low price or fewer ties at a higher price. Table 12—2 is a reproduction of Table 10—3A, but with the addition of data on marginal costs, marginal revenues, marginal products, and the sale values of these marginal products.

At any input short of 8, and an associated output short of 2,908, this tie manufacturer would be failing to make the most of his opportunities.

TABLE 12—2

THE COSTS AND REVENUES OF A HYPOTHETICAL MANUFACTURER OF TIES

(BUYING PURELY COMPETITIVELY; SELLING MONOPOLISTICALLY)

<i>Input</i>	<i>Output</i>	<i>Average revenue</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Average expenditure</i>	<i>Marginal cost</i>	<i>Marginal revenue</i>	<i>Marginal product</i>	<i>Sale value of marginal product</i>
1	100	\$2.80	\$1,100	\$ 280.00	\$ 820.00	\$100	\$100	\$ 280.00	100	\$ 280.00
2	350	2.75	1,200	962.50	-237.50	100	100	682.50	250	687.50
3	702	2.65	1,300	1,860.30	560.30	100	100	897.80	352	932.80
4	1,152	2.50	1,400	2,880.00	1,480.00	100	100	1,019.70	450	1,125.00
5	1,700	2.30	1,500	3,910.00	2,410.00	100	100	1,030.00	548	1,260.40
6	2,190	2.10	1,600	4,599.00	2,999.00	100	100	689.00	490	1,029.00
7	2,604	1.90	1,700	4,947.60	3,247.60	100	100	348.60	414	786.60
8	2,908	1.75	1,800	5,089.00	3,289.00	100	100	141.40	304	532.00
9	3,114	1.65	1,900	5,138.10	3,238.10	100	100	49.10	206	339.90
10	3,240	1.55	2,000	5,022.00	3,022.00	100	100	-116.10	126	195.30
11	3,300	1.45	2,100	4,785.00	2,685.00	100	100	-237.00	60	87.00
12	3,350	1.35	2,200	4,522.50	2,322.50	100	100	-262.50	50	67.50
13	3,395	1.25	2,300	4,243.75	1,943.75	100	100	-278.75	45	56.25
14	3,435	1.15	2,400	3,950.25	1,550.25	100	100	-293.50	40	46.00

Expansion would add more to revenues than to costs; that is marginal revenues from the addition of the sixth and the seventh and the eighth units of input (with the associated increases in outputs) are in each case greater than the marginal costs involved. At any production level greater than 8 inputs producing 2,908 ties, this manufacturer would be making less profits than would be possible. The addition of the ninth and the tenth units of input (with the associated increases in outputs) adds more to costs than to revenues; that is, in these ranges marginal cost is greater than marginal revenue. Marginal analysis tells us that his most profitable level of production will be at 8 units of input turning out 2,908 ties. This again is the same conclusion which we would have reached by direct comparisons between total revenues and total costs, as in Chapter 10.

In maximizing his profits this tie manufacturer is not acting consistently with the interests of consumers. He is curtailing output in order to get a higher price. For example the ninth unit of input would add only \$100 to costs, while the sale value of the marginal product which would result from such an expansion is \$339.90. *He fails to continue to expand so long as the addition to his costs is less than the worth to consumers of the added product.* The adjustment most consistent with response to consumer preferences would be at a production level employing 10 units of input and producing 3,240 ties, these ties being sold at \$1.55 each. But at such a production level this entrepreneur's profits would be less than when he restricts operations to turn out only 2,908 ties sold at \$1.75 apiece. His motives are the same as those of the purely competitive seller; but his behavior is not consistent with consumer welfare, while that of the purely competitive seller is.

A hypothetical firm that buys monopsonistically and sells purely competitively

Marginal analysis of the most profitable level of production of a hypothetical clock manufacturing enterprise such as was described in Chapter 10 again follows the same broad pattern as in the examples of the hose and tie firms. We arrive at the most profitable adjustment by comparing marginal revenues with marginal costs. The clock producer was assumed to sell clocks in national markets in pure competition with many other manufacturers of clocks exactly like his; he was assumed to buy skilled labor in a geographically somewhat isolated local labor market. In order to simplify the analysis we made the further assumption that this labor was the only variable agent, and this assumption will be continued though it is not necessary to the conclusions which we shall reach. Table 12—3 is a reproduction of Table 10—4, but with the addition of data on marginal costs, marginal revenues, marginal products, and the sale values of these marginal products.

At any input of skilled clock makers short of 9, and associated output of clocks short of 3,114, this entrepreneur would be failing to make the most of his opportunities. Expansion would add more to revenues than

TABLE 12—3

THE COSTS AND REVENUES OF A HYPOTHETICAL MANUFACTURER OF CLOCKS

(BUYING MONOPSONISTICALLY; SELLING PURELY COMPETITIVELY)

<i>Input</i>	<i>Output</i>	<i>Average revenue</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Average expenditure</i>	<i>Marginal cost</i>	<i>Marginal revenue</i>	<i>Marginal product</i>	<i>Sale value of marginal product</i>
1	100	\$1.80	\$1,075	\$ 180.00	\$ -895.00	\$ 75	..	\$180.00	100	\$180.00
2	350	1.80	1,170	630.00	-540.00	85	95	450.00	250	450.00
3	702	1.80	1,285	1,263.60	-21.40	95	115	633.60	352	633.60
4	1,152	1.80	1,420	2,073.40	653.40	105	135	810.00	450	810.00
5	1,700	1.80	1,575	3,060.00	1,485.00	115	155	986.40	548	986.40
6	2,190	1.80	1,760	3,942.00	2,192.00	125	175	882.00	490	882.00
7	2,604	1.80	1,945	4,687.20	2,742.20	135	195	745.20	414	745.20
8	2,908	1.80	2,160	5,234.40	3,074.40	145	215	547.20	304	547.20
9	3,114	1.80	2,395	5,605.20	3,210.20	155	235	370.80	206	370.80
10	3,240	1.80	2,650	5,832.00	3,182.00	165	255	226.80	126	226.80
11	3,300	1.80	2,925	5,940.00	3,015.00	175	275	108.00	60	108.00
12	3,350	1.80	3,220	6,030.00	2,810.00	185	295	90.00	50	90.00
13	3,395	1.80	3,535	6,111.00	2,576.00	195	315	81.00	45	81.00
14	3,435	1.80	3,870	6,183.00	2,313.00	205	335	72.00	40	72.00

to costs; that is marginal revenues from the addition of the eighth or the ninth workers to input (with the associated increases in outputs) are in each case greater than the marginal costs involved. At any production level greater than 9 inputs of skilled clock makers, producing 3,114 clocks, this manufacturer would be making less profits than would be possible. The addition of the tenth and the eleventh units of input (with the associated increases in outputs) adds more to costs than to revenues. Marginal analysis tells us that his most profitable level of production will be at 9 units of input, turning out 3,114 clocks. This again is the same conclusion that we would have reached by direct comparisons between total revenues and total costs.

In maximizing his profits this clock manufacturer is not acting consistently with the interests of consumers (or of skilled clock makers). He is curtailing inputs, and hence outputs, in order to pay lower wages (average expenditures). It would take only \$165 to bid the tenth clock maker away from other occupations or other localities; his worth in these other positions is only \$165. The sale value of the marginal product he would add here is \$227. *This clock manufacturer fails to continue to expand his level of production so long as the worth of another unit of input in other uses is less than the worth of the contribution it would make here.* The adjustment most consistent with consumer preferences (and with increasing the opportunities for workers who might become, or already are, skilled clock makers) would be at 10 units of input, paying wages of \$165, and turning out 3,240 clocks. Again, the motives of the clock manufacturer are assumed to be the same as those of the hose and the tie manufacturers; all are seeking to maximize profits. But the behavior of the monopsonistic buyer is inconsistent with consumer interests in the allocating of resources to clock making; he is curtailing production of clocks when consumers would rather see more labor services used in this way even at the sacrifice of some labor services used in making other things.

A hypothetical firm that buys monopsonistically and sells monopolistically

Finally, we may turn to our fourth acquaintance, the entrepreneur of a box factory that is buying the services of skilled box carvers monopsonistically and selling the boxes in monopolistic markets. These carvers are assumed to be the only variable agents employed in the factory; a unit of input is therefore identical with the input of the services of one box carver for the period under consideration. We again follow the same procedure of comparing marginal revenues with marginal costs. Table 12—4 is a reproduction of Table 10—5, but with the addition of data on marginal revenues, marginal costs, marginal products, and the sale values of these marginal products.

At any input of skilled carvers short of 7, and associated output of boxes short of 2,604, this entrepreneur would be failing to make the most of his opportunities. Expansion would add more to revenues than to

TABLE 12—4

THE COSTS AND REVENUES OF A HYPOTHETICAL MANUFACTURER OF BOXES
(BUYING MONOPSONISTICALLY; SELLING MONOPOLISTICALLY)

<i>Input</i>	<i>Output</i>	<i>Average revenue</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Average expenditure</i>	<i>Marginal cost</i>	<i>Marginal revenue</i>	<i>Marginal product</i>	<i>Sale value of marginal product</i>
1	100	\$2.80	\$1,075	\$ 280.00	\$ -795.00	\$ 75	\$. .	\$ 280.00	100	\$ 280.00
2	350	2.75	1,170	962.50	-207.50	85	95	682.50	250	687.50
3	702	2.65	1,285	1,860.30	575.30	95	115	897.80	352	932.80
4	1,152	2.50	1,420	2,880.00	1,460.00	105	135	1,019.70	450	1,125.00
5	1,700	2.30	1,575	3,910.00	2,335.00	115	155	1,030.00	548	1,260.40
6	2,190	2.10	1,750	4,599.00	2,849.00	125	175	689.00	490	1,029.00
7	2,604	1.90	1,945	4,947.60	3,002.60	135	195	348.60	414	786.60
8	2,908	1.75	2,160	5,089.00	2,929.00	145	215	141.40	304	532.00
9	3,114	1.65	2,395	5,138.10	2,743.10	155	235	49.10	206	339.90
10	3,240	1.55	2,650	5,022.20	2,372.00	165	255	-116.10	126	195.30
11	3,300	1.45	2,925	4,785.00	1,860.00	175	275	-237.00	60	87.00
12	3,350	1.35	3,220	4,522.50	1,302.50	185	295	-262.50	50	67.50
13	3,395	1.25	3,535	4,243.75	708.75	195	315	-278.75	45	56.25
14	3,435	1.15	3,870	3,950.25	80.25	205	335	-293.50	40	46.00

costs; marginal revenues from the addition of the sixth or the seventh workers to input (with the associated increase in output of boxes) are in each case greater than the marginal costs involved. At any production level greater than 7 inputs of skilled box carvers (producing 2,604 boxes) this manufacturer would be making less profits than would be possible. The addition of the eighth and ninth units of input (with the associated increases in outputs) adds more to costs than to revenues; that is, in these ranges marginal costs are greater than marginal revenues. Marginal analysis tells us that his most profitable level of production will be employing 7 box carvers, turning out 2,604 boxes. Once more our conclusion arrived at through marginal analysis is the same as that that we would reach by direct comparisons between total revenues and total costs.

In maximizing his profits the manufacturer of boxes is acting inconsistently with consumer preferences for two reasons: He is curtailing production in order to charge a higher price for his boxes, and he is curtailing production in order to pay lower wages to carvers. The sale values of marginal products are greater than marginal revenues; and the average expenditures are less than marginal costs. He therefore fails to employ workers whose worth elsewhere is less than the marginal revenues they would bring to him; and he fails to expand outputs when the addition to his costs would be less than the worth of the added product to consumers. The "ideal" level of production would be with the employment of 10 units of input, paying wages of \$165 and turning out 3,240 boxes for sale at a price of \$1.55 each; but this would not bring the box manufacturer profits as great as he can get by curtailing production to an input of 7 paying wages of \$135 and turning out 2,604 boxes for sale at \$1.90 each. Again he is acting, like other entrepreneurs, to maximize his profits; but in so acting he is behaving inconsistently with the interests of consumers (and of workers) in both the market in which he sells and the market in which he buys.

Graphic Representation of Marginal Input Adjustments

The contrast between purely competitive situations and those in which nonopolistic or monopsonistic conditions are present is brought out vividly by the use of graphs. Figure 12—1 gives a set of smoothed graphs for hypothetical situations combining different sets of conditions on the buying and the selling sides of the firm's contact with the market. The horizontal axis on these graphs is marked off by units of input. The curves indicate the marginal costs, average expenditures, marginal revenues, and sale values of the marginal products at each input level. At an input of 100 units, for example, the marginal cost curve indicates the difference between total costs when 99 and when 100 units of input are employed.

Figure 12—1A represents a purely competitive firm. One curve (MC ; AE) represents both marginal cost and average expenditure, and since average expenditure does not change with changes in input this

curve appears as a horizontal line. The other curve (MR ; SV) represents both marginal revenue and the sale values of the marginal products. If it were possible to add inputs a fraction at a time, this firm would maximize profits by adjusting to the point at which the two curves intersect; any point short of that indicates that additions to cost would be less than additions to revenue, whereas beyond the intersection the

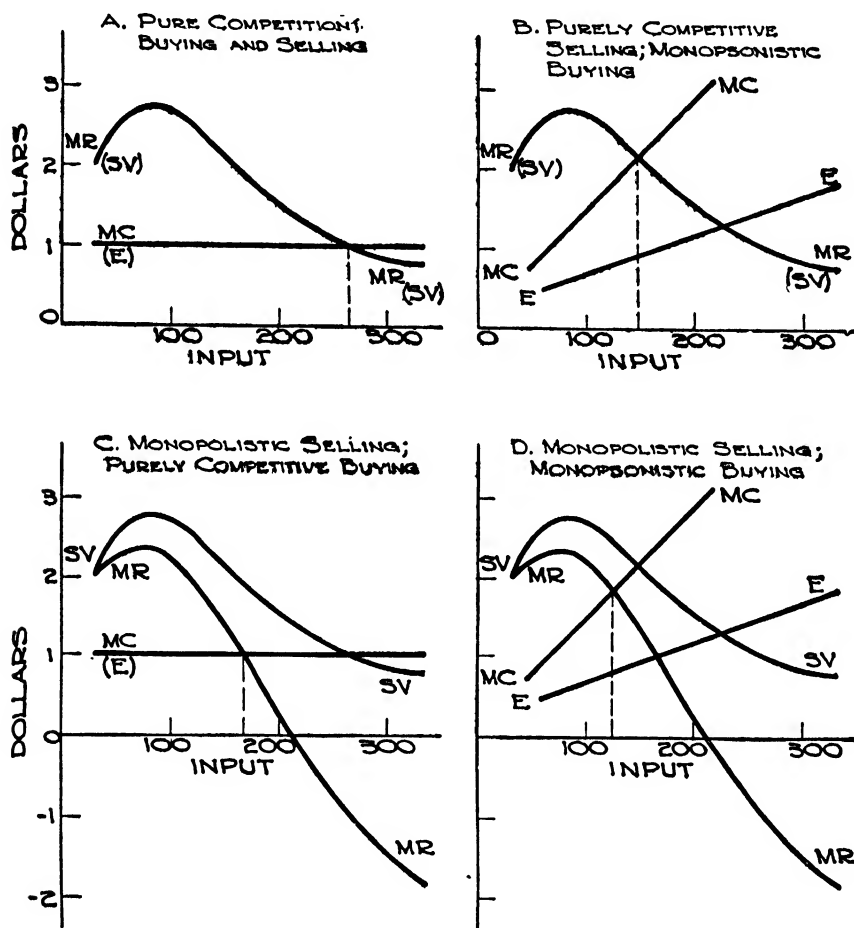


Fig. 12—1. Input adjustments.

reverse situation appears. This most profitable, purely competitive adjustment is also that which is consistent with allocating resources in accordance with consumer preferences. So long as the worth to consumers of the product added by another unit of input (the sale value of the marginal product) is greater than the worth of another unit of input in alternative uses (the average expenditure) production is expanded.

Figure 12—1D is the other extreme, of the monopsonistic buyer selling monopolistically. Assuming inputs can be varied a fraction at a time,

his most profitable adjustment would be where the marginal cost and the marginal revenue curves intersect. This is short of the point where average expenditure equals the sale value of the marginal product for two reasons: because he is curtailing production in order to raise the price of what he sells, and because he is curtailing production in order to lower the price of what he buys. For both reasons his behavior is inconsistent with the allocating of resources in accordance with consumer preferences. The "ideal" adjustment from this point of view would be where the average expenditure and the sale value of marginal product curves intersect. A comparison between the "ideal" and the most profitable adjustments gives a clear visual evidence of the curtailment involved in the monopolistic-monopsonistic position of the firm.

Figures 12—1B and 12—1C are intermediate cases. Curtailment in the case of the monopolistic seller who buys productive services in purely competitive markets is illustrated in Figure 12—1C; he cuts down his level of production in order to boost his price (marginal revenue is less than the sale value of the marginal product). Curtailment in the case of the monopsonistic buyer who sells his product in a purely competitive market is illustrated in Figure 12—1B; he cuts down his level of production in order to lower the prices he pays for productive services (marginal cost is greater than average expenditure). The difference between the ideal and the most profitable adjustments again show up clearly.

CHAPTER 13

Analysis of Marginal Adjustments in Terms of Outputs

THUS far our marginal analysis has been focused on the shifts which occur in the inputs of firms, while the associated outputs were only indirectly considered. But frequently it is important to have a means of focusing on the output side of the adjustment. In looking primarily at inputs we emphasize the allocation of resources to one use or another; in looking at outputs we study the same thing, but with a focus on what it is that ultimately is likely to get produced and in what quantities. We may want to know, for example, how much will be offered for sale by purely competitive sellers of raw cotton at various market prices for cotton, or by monopolistically competitive sellers of canned peas at various general levels of prices for peas. We may be interested in the effects of a tax of 5 cents on admissions to movies or of 3 cents per package of cigarets. These questions could be answered by the use of the analysis of Chapter 10 and the marginal analysis as presented in Chapter 12, that is by making comparisons between marginal revenues and marginal costs; but such questions are more easily answered by the use of another kind of marginal analysis, by comparing the addition to cost *per unit added to output* with the addition to revenue *per unit added to output*. The first of these is called *marginal unit cost*, and the second is called *marginal unit revenue*.

Examination of Concepts

Deriving figures for marginal unit cost and marginal unit revenue from discontinuous production data

If outputs could be increased or decreased one unit at a time, then finding marginal unit cost and marginal unit revenue would be a very simple matter. The marginal unit cost at the one thousand and first can of peas would be the difference between the total cost of producing 1,000 cans of peas and 1,001 cans of peas; the marginal unit revenue at the one thousand and first can of peas would be the difference between the total revenue from selling 1,000 cans of peas and 1,001 cans of peas. Unfortunately for students of economics, production adjustments are not made in that way. The manufacturer of candy bars does not ask whether

he would gain by adding one more candy bar when he is producing 63,257 bars; rather he considers whether hiring another worker and buying another case of raw chocolate would be advantageous, and this would probably increase his output of candy bars from, say, 63,000 to 63,500. Production adjustments are *discontinuous*, going in jumps. Marginal unit cost is then the addition to cost *per bar added* in making this expansion (from 63,000 to 63,500 units of output). Marginal unit revenue is the addition to revenue *per bar added* in this adjustment. Estimates of marginal unit cost and of marginal unit revenue are then found in this way:

1. *Marginal unit cost* is marginal cost divided by marginal product:

$$MUC = \frac{MC}{MP}$$

2. *Marginal unit revenue* is marginal revenue divided by marginal product: $MUR = \frac{MR}{MP}$

This relationship may be clarified by drawing once more on the same extremely simple table we used in defining marginal revenue and marginal cost in Chapter 11. We assumed a puzzle maker to have costs and revenues as given in Table 11—1 of that chapter, and as repeated in Table 13—1 here. We have added to this table data on marginal unit costs and on marginal unit revenues.

TABLE 13—1

HYPOTHETICAL PUZZLE-MANUFACTURING FIRM (CONTINUED)

<i>Input</i>	<i>Total product</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Marginal product</i>	<i>Marginal cost</i>	<i>Marginal revenue</i>	<i>Marginal unit cost</i>	<i>Marginal unit revenue</i>
4	1,000	\$400	\$1,000	?	\$?	\$?	\$?	\$?
5	1,200	525	1,140	200	125	+140	.63	+.70
6	1,350	660	1,215	150	135	+75	.90	+.70
7	1,450	805	1,160	100	145	-55	1.45	-.55

When the puzzle manufacturer added the fifth unit of input he increased his costs by \$125 (marginal cost), which amounted to \$.63 per puzzle added to output. This \$.63 is found by dividing the marginal cost of \$125 by the marginal product of 200 puzzles. The expanded sales resulting from adding the fifth unit to input resulted in an increase of \$140 in revenue (marginal revenue), which amounted to \$.70 per puzzle added to output, and hence to sales. This \$.70 is found by dividing \$140 by the marginal product of 200 puzzles. The other figures for marginal unit cost and marginal unit revenue are derived in the same way. When marginal revenue is negative, marginal unit revenue will of course be negative also.

The relation between marginal unit revenue and average revenue

Under purely competitive conditions marginal unit revenue and average revenue will be equal. This occurs for the same reason that under purely

competitive conditions, marginal revenue and the sale value of the marginal product are equal. *When a purely competitive firm expands its sales it adds to revenue exactly the price of the additional units times the number of units added. No lowering of price is involved.*

Under monopolistic selling conditions, marginal unit revenue will generally be less than average revenue. This occurs for the same reason that under monopolistic conditions marginal revenue is less than the sale value of the marginal product. When a monopolistic firm expands its sales it must lower the price charged; therefore the reduction of revenue received per unit of sales (average revenue) is set off against the addition to revenue which comes from selling more units.

The relations between marginal unit revenues and average revenues in different market situations may be clarified by some very simple algebra:

Let AR = Average revenue

MUR = Marginal unit revenue

SV = Sale value of marginal product

MR = Marginal revenue

MP = Marginal product

By definition: (1) $MUR = \frac{MR}{MP}$

(2) $SV = AR \times MP$; therefore $AR = \frac{SV}{MP}$

Under purely competitive conditions:

$$SV = MR$$

$$\therefore \frac{SV}{MP} = \frac{MR}{MP}, \text{ and } AR = MUR$$

Under monopolistic conditions:

$$SV > MR$$

$$\therefore \frac{SV}{MP} > \frac{MR}{MP}, \text{ and } AR > MUR$$

If outputs could be expanded continuously, one unit at a time, marginal product would be one. In such a case, $\frac{SV}{MP} = SV$, and $\frac{MR}{MP} = MR$; in such a case SV is identical with AR , and MR is identical with MUR .

Just as the relation between marginal revenue and the sale value of marginal product were shown graphically in terms of units of input, so marginal unit revenue and average revenue may be plotted on a graph marked off by *outputs*. This has been done in Figure 13—1. Under purely competitive conditions the average revenue and marginal unit revenue curves coincide in a horizontal line at the market price. Under monopolistic conditions, marginal unit revenue falls consistently below average revenue, and both slope down to the right.

Failure of marginal output analysis to discriminate between pure competition and monopsony in buying

This output-oriented comparison between marginal unit revenue and average revenue (price) on the selling side is simply another way of looking at the relation between marginal revenue and the sale value of

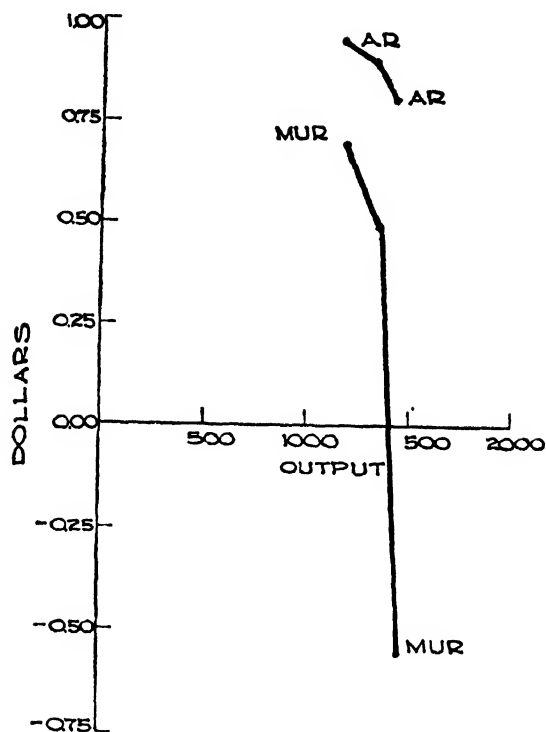


Fig. 13—1. Average revenue and marginal unit revenue of a hypothetical firm.

the marginal product. But the analogous comparison between marginal cost and average expenditure on the buying side cannot be expressed in terms of outputs. There is no *direct* way of comparing marginal unit costs with the prices of units of input (average expenditure), since marginal unit cost is expressed in terms of units of output while average expenditure is expressed in units of input. Analysis by the use of marginal unit cost and marginal unit revenue figures therefore fails to discriminate between purely competitive and monopsonistic situations in buying. Instead of having four sets of conditions to consider in applying this marginal output analysis, we have only two: purely competitive sellers (whose buying positions are not distinguished) and monopolistic sellers (whose buying positions are not distinguished).

Finding the Most Profitable (or Least Unprofitable) Output by Comparing Marginal Unit Revenues and Marginal Unit Costs

In order to illustrate the nature of the comparison between marginal unit revenues and marginal unit costs in arriving at conclusions concerning the most profitable adjustment of production, we shall call once more on the hypothetical entrepreneurs who have served us in preceding

chapters. We might illustrate the purely competitive output adjustment by taking either the hose manufacturer or the clock manufacturer; we shall choose the hose manufacturer. We might illustrate the monopolistic output adjustment by taking either the tie manufacturer or the entrepreneur of the box-making firm; we shall choose the tie manufacturer. Data for these firms are once more reproduced in Tables 13—2 and 13—3, with figures added for marginal unit revenues and marginal unit costs.

By running down the table representing the purely competitive hose manufacturer (Table 13—2) we find that marginal unit revenue exceeds marginal unit cost at all outputs up to and including 3,300 pairs of hose (with an input of 11); it is less than marginal unit cost at greater outputs than this. Again we arrive at the conclusion that the most profitable output for the hose producer is 3,300 pairs (produced by 11 units of input). This is not surprising. Since marginal unit revenue is marginal revenue divided by marginal product, and marginal unit cost is marginal cost divided by the same marginal product, it must follow that whenever marginal revenue exceeds marginal cost, marginal unit revenue will exceed marginal unit cost. Marginal unit revenue is the amount by which revenue is increased per unit added to output, and marginal unit cost is the amount by which cost is increased per unit added to output. If by expanding the entrepreneur can increase revenue per unit added to output by more than he increases cost per unit added to output, he would clearly be able to improve his position by shifting to the higher level of production. On the other hand expansion would be unwise if it would increase cost per unit added to output more than it would increase revenue per unit added to output.

By running down the table representing the monopolistic seller of ties (Table 13—3) we find that marginal unit revenue exceeds marginal unit cost at all outputs up to and including 2,908 ties (with an input of 8); it is less than marginal unit cost beyond this point. Again our conclusions check with those arrived at in previous examinations of this firm.

We may now note an important difference between these two firms. In the case of the purely competitive seller marginal unit revenue equaled average revenue, or selling price; *we could therefore have found the most profitable adjustment for the purely competitive seller simply by comparing market price with marginal unit costs at various output levels.* The purely competitive seller would expand output so long as marginal unit cost was less than selling price. The importance of this fact will become more evident as we proceed with studies of market adjustments. Here it is sufficient to say that this is the counterpart of the relation between marginal cost and the sale value of the marginal product, and it indicates behavior consistent with consumer interests in the selling side of the firm's activities. In the case of the monopolistic seller, on the other hand, marginal unit revenue is always less than average revenue, and the firm therefore curtails output when further expansions would still involve marginal unit costs less than the price at which the greater output could be sold.

TABLE 13--2

COST AND REVENUE FOR A HYPOTHETICAL FIRM SELLING PURELY COMPETITIVELY

<i>Input</i>	<i>Output</i>	<i>Average revenue</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Marginal product</i>	<i>Marginal cost</i>	<i>Marginal unit cost</i>	<i>Marginal revenue</i>	<i>Marginal unit revenue</i>
1	100	\$1.80	\$1,100	\$ 180.00	\$ -920.00	100	\$100	\$1.00	\$180.00	\$1.80
2	350	1.80	1,200	630.00	-570.00	250	100	.40	450.00	1.80
3	702	1.80	1,300	1,263.60	-36.40	352	100	.28	633.60	1.80
4	1,152	1.80	1,400	2,073.60	673.60	450	100	.22	810.00	1.80
5	1,700	1.80	1,500	3,060.00	1,560.00	548	100	.18	986.40	1.80
6	2,190	1.80	1,600	3,942.00	2,342.00	490	100	.20	882.00	1.80
7	2,604	1.80	1,700	4,687.20	2,987.20	414	100	.24	745.20	1.80
8	2,908	1.80	1,800	5,234.40	3,434.40	304	100	.33	547.20	1.80
9	3,114	1.80	1,900	5,605.20	3,705.20	206	100	.49	370.80	1.80
10	3,240	1.80	2,000	5,832.00	3,832.00	126	100	.79	226.80	1.80
11	3,300	1.80	2,100	5,940.00	3,840.00	60	100	1.67	108.00	1.80
12	3,350	1.80	2,200	6,030.00	3,830.00	50	100	2.00	90.00	1.80
13	3,395	1.80	2,300	6,111.00	3,811.00	45	100	2.22	81.00	1.80
14	3,435	1.80	2,400	6,183.00	3,783.00	40	100	2.50	72.00	1.80

TABLE 13—3
COST AND REVENUE FOR A HYPOTHETICAL FIRM SELLING MONOPOLISTICALLY

<i>Input</i>	<i>Output</i>	<i>Average revenue</i>	<i>Total cost</i>	<i>Total revenue</i>	<i>Profit</i>	<i>Marginal product</i>	<i>Marginal cost</i>	<i>Marginal unit cost</i>	<i>Marginal revenue</i>	<i>Marginal unit revenue</i>
1	100	\$2.80	\$1,100	\$ 280.00	\$-820.00	100	\$100	\$1.00	\$ 280.00	\$2.80
2	350	2.75	1,200	962.50	-237.50	250	100	.40	682.50	2.73
3	702	2.65	1,300	1,860.30	560.30	352	100	.28	897.80	2.55
4	1,152	2.50	1,400	2,880.00	1,480.00	450	100	.22	1,019.70	2.27
5	1,700	2.30	1,500	3,910.00	2,410.00	548	100	.18	1,030.00	1.88
6	2,190	2.10	1,600	4,599.00	2,999.00	640	100	.20	689.00	1.41
7	2,604	1.90	1,700	4,947.60	3,247.60	714	100	.24	348.60	.84
8	2,908	1.75	1,800	5,089.00	3,289.00	764	100	.33	141.40	.47
9	3,114	1.65	1,900	5,138.10	3,238.10	806	100	.49	49.10	.24
10	3,240	1.55	2,000	5,022.00	3,022.00	840	100	.79	-116.10	-.91
11	3,300	1.45	2,100	4,785.00	2,685.00	860	100	1.67	-237.00	-3.95
12	3,350	1.35	2,200	4,522.50	2,322.50	870	100	2.00	-262.50	-5.25
13	3,395	1.25	2,300	4,243.75	1,943.75	875	100	2.22	-278.75	-6.19
14	3,435	1.15	2,400	3,950.25	1,550.25	880	100	2.50	-293.50	-7.34

The Relations Between Marginal Unit Costs and Other Cost Data

By making the simplifying assumption that outputs can be varied continuously (one unit or even fractions of a unit at a time), it is easy to show the relation between marginal unit costs and total unit costs and the relation between marginal unit costs and variable unit costs.

The relation between marginal unit costs and variable unit costs

Variable unit cost is simply an average; it is the variable cost at a given output divided by the number of units in that output. This concept was presented in some detail in Chapter 7 and was used in Chapter 10. Suppose now that output is increased by one unit; the addition to cost involved is then marginal unit cost: it is an addition to variable costs. If this addition to variable cost is greater than the previous average (the previous variable unit cost) it will pull that average up. If this addition to variable cost is less than the previous average, it will pull that average down. *Thus whenever marginal unit cost is above variable unit cost, variable unit cost must rise; whenever marginal unit cost is below variable unit cost, variable unit cost must fall.* This is exactly what happens in the firms represented by the data of Tables 13—2 and 13—3.

The relation between marginal unit cost and total unit cost

Total unit cost is also an average; it is the total cost at a given output divided by the number of units in that output. When something is added to cost, it of course increases total cost, just as it increases variable cost. Suppose again that output is increased by one unit. If the addition to cost resulting (marginal unit cost) is greater than the previous average (total unit cost) it will pull that average up; if it is less it will pull that average down. *Thus whenever marginal unit cost is above total unit cost, total unit cost must rise; whenever marginal unit cost is below total unit cost, total unit cost must fall.* This relation is also illustrated by the data of Tables 13—2 and 13—3.

Cost-Output Curve Analysis and Short-Run Adjustments of the Firm

Presentation of cost and revenue curves

Graphic analysis representing output adjustments of the firm in the short run was presented in some detail in Chapter 10. There was one major gap in that graphic analysis, however. There was no way of showing clearly on the graph what the most profitable output would be; we simply had to read that from the table by looking for the level at which profits would be greatest (or losses least) and then apply our knowledge to construction of the graph. Now that we have introduced the concepts of marginal unit cost and marginal unit revenue we have filled that gap in our graphic presentation. By drawing in curves to

represent marginal unit costs and marginal unit revenues at various outputs we can determine graphically what the most profitable, or the least unprofitable, output will be. So long as marginal unit revenue is greater than marginal unit cost, it will pay to expand; but if marginal unit revenue is less than marginal unit cost, then curtailment of output would improve the position of the firm.

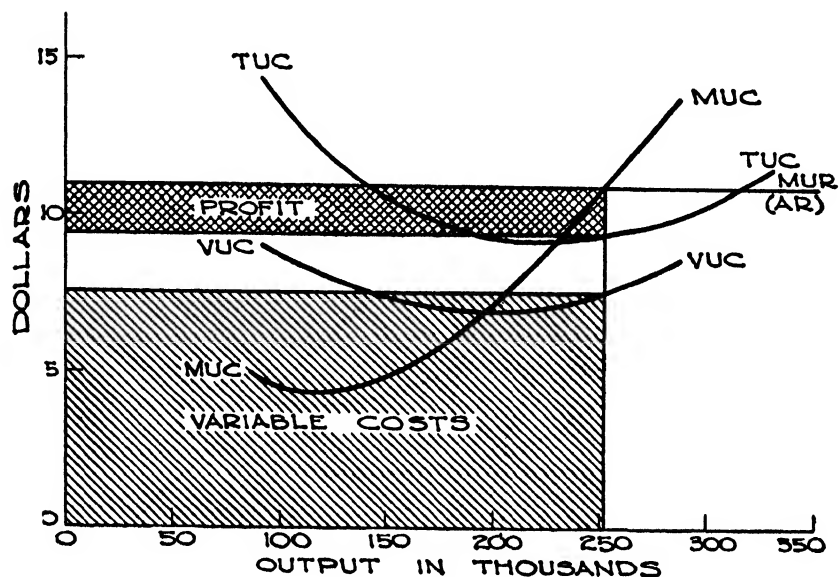


Fig. 13—2A. Purely competitive firm operates at a profit.

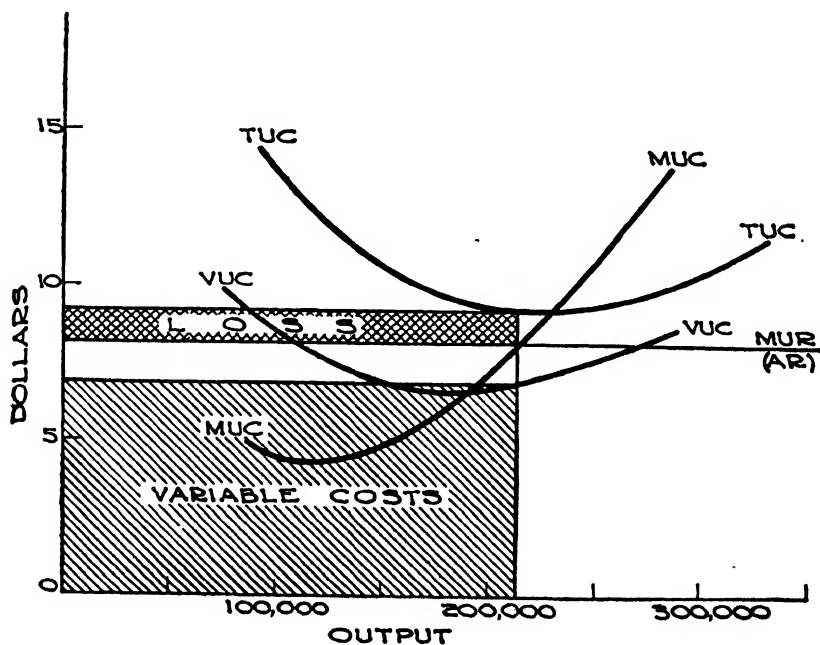


Fig. 13—2B. Purely competitive firm operates at a loss.

202 MARGINAL ANALYSIS IN TERMS OF OUTPUTS

If we make the simplifying assumption that outputs can be varied one unit (or a fraction of a unit) at a time, the curves are smooth. Moreover, we can then find the most profitable output level from the point of intersection of the marginal unit revenue and the marginal unit cost curves. At any output short of this point marginal unit revenue is greater than marginal unit cost, and the firm could improve its position by expanding; at any output greater than this, marginal unit revenue

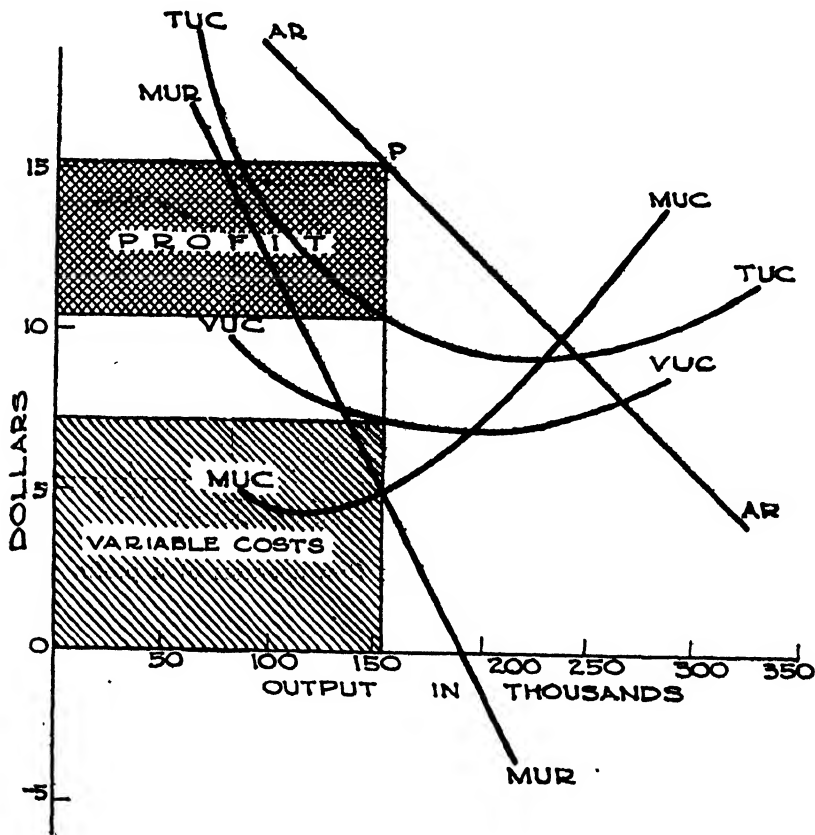


Fig. 13—3A. Monopolistic firm operates at a profit.

is less than marginal unit cost, and the firm could improve its position by reducing its output.

Here we have drawn some hypothetical cost and revenue curves expressed in relation to units of output, assuming that it is possible to vary outputs by one unit (or by fractions of a unit) at a time. The marginal unit cost curves cut the total unit cost curves at their lowest points, and they cut the variable unit cost curves at their lowest points. This must be true if the graphs are drawn correctly, because of the relations described above. The average revenue curve for a purely competitive seller is a horizontal line, indicating that whatever his output he will get the going market price; and since marginal unit revenue and average

revenue coincide under purely competitive selling conditions, there is only one revenue line in the graphs representing the purely competitive firms.

There are, however, two separate revenue curves for the monopolist; marginal unit revenues are below average revenues at all output levels.

Figure 13—2A and Figure 13—2B represent purely competitive sellers; the first can make a profit, the second will operate at a loss. These figures are analogous to Figures 10—1, 10—2, and 10—3; and for an explana-

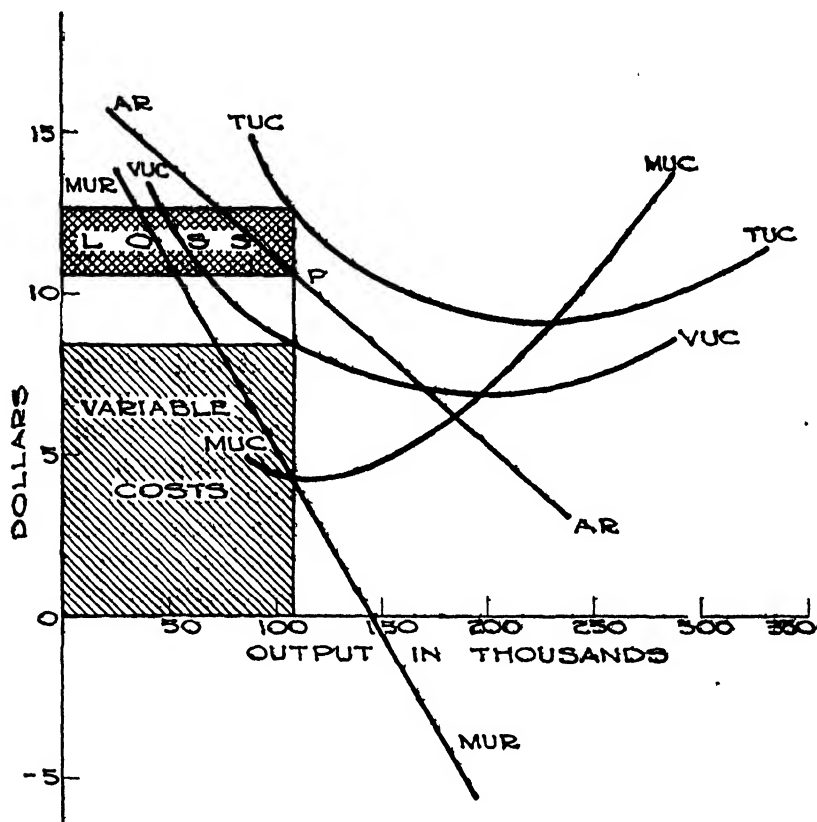


Fig. 13—3B. Monopolistic firm operates at a loss.

tion of them the reader may turn back to that analysis. The only additional feature, albeit an important one, is the addition of the marginal unit cost curve which makes it possible to determine graphically what will be the most profitable (or the least unprofitable) output. The intersection of this curve with the average revenue-marginal unit revenue line indicates the most profitable output level.

Figures 13—3A and 13—3B represent monopolistic sellers; the first can make a profit, the second will operate at a loss. These figures are analogous to Figures 10—4 and 10—5. Here two new curves are added, marginal unit revenues and marginal unit costs; the most profitable output is found graphically by determining the point of intersection of these curves and then reading from the graph the output at which the

intersection appears. At any output less than this the firm could add more to revenue than to cost by expanding; but increases of output beyond this point would add more to costs than to revenues. Again, interpretation of the other features of the graph may be facilitated by a review of the discussion in Chapter 10.

Maximum profit (or minimum loss) output and the least-cost-combination

Figure 13—2A illustrates the relation between the maximum profit output and the least-cost-combination, which was remarked upon in Chapter 10. Let us assume outputs to be continuously variable. Then whenever selling price is above lowest total unit cost in a purely competitive firm, so that profits are obtainable, the output adjustment *must* be beyond the least-cost-combination point; this is true because a purely competitive entrepreneur will continue expanding production until marginal unit cost equals average revenue (or selling price). When marginal unit cost is below total unit cost it pulls total unit cost down; when marginal unit cost is above total unit cost, total unit cost will rise. Therefore the point at which marginal unit cost and selling price are equal in a *profitable* purely competitive firm must be beyond the upward turn of total unit cost; it is only in this range that marginal unit cost will rise high enough to equal the high level of selling price.

Figure 13—2B illustrates the relation between minimum loss output and the least-cost-combination, which was also noted in Chapter 10. Again we assume outputs to be continuously variable. Whenever selling price is below the lowest total unit cost in a purely competitive firm, the output adjustment *must* be short of the least-cost-combination point. So long as marginal unit cost is less than total unit cost, total unit cost will be falling; and it is clearly only in the range in which marginal unit cost is less than total unit cost that it will equal a selling price less than the lowest total unit cost. In attempting to minimize losses the entrepreneur will therefore find that his most favorable position, where marginal unit cost is equal to average revenue (selling price), is short of the least-cost-combination point.

In a profitable monopolistic firm it is entirely possible that marginal unit cost and marginal unit revenue might intersect *either* before or beyond the least-cost-combination point. In the case of a monopolistic seller operating at a loss, however, the least unfavorable output adjustment will inevitably be short of the least-cost-combination position. It is only when there is no output at which average revenue is above the associated total unit cost that a monopolistic seller would be unable to make a profit. If an average revenue curve lies below the total unit cost curve at all points, the associated marginal unit revenue curve must lie even lower; marginal unit cost and marginal unit revenue must then intersect at an output level short of the least-cost-combination output.¹

¹ Assuming a downward sloping average revenue curve.

Part III

PURELY COMPETITIVE MARKET ADJUSTMENTS

PREAMBLE

IF THE entire economy were purely competitive, the study of economics would be very simple. But the economy is not that way. Then why do we not just forget about pure competition and start out with the way the economy *really* is? The answer to that question is very simple: the economy is too complex. We do not want to spend our lives on a host of details about economic matters and come out with a long list of infinitesimal facts that do not help us a whit in understanding the effects of excise taxes and tariffs and work relief and wage legislation and a lot of other important things. Pure competition is just one of the simplified models that help us trace a thread of understanding of economic adjustments. It is an especially satisfying model, because it gives us clues about the whole of the economy, whereas our other models come in smaller pieces; it is also an especially satisfying model because if an economy were to be completely competitive it would do remarkably efficiently the job of giving consumers the things they prefer. It is a dangerous model if we forget that it gives only a partial and inaccurate picture of actual economic adjustments.

There are two reasons, then, for doing a lot of talking about pure competition. First, in many respects the economy is *almost* like what it would be if it were purely competitive; and it is a lot easier to start by describing pure competition than to describe the economy the way it actually is. We draw a map giving some important points; then later we put other things on the map, such as monopsony and monopoly and oligopoly and oligopsony and quality competition. Second, consumers would come out remarkably well in a purely competitive economy; in that respect it provides a useful model with which we can compare some of the things that happen in other kinds of situations. It gives us ideas as to what we might do about some of the things we don't like in the way the economy actually works. Especially it gives us ideas about possibilities of raising planes of living and getting produced the things we want most, instead of the things we do not want so much, or even (sometimes) instead of nothing at all.

CHAPTER 14

Pure Competition and the Law of Supply and Demand

THERE is something almost magical about the way in which market price is determined in a purely competitive industry. Nobody seems to have anything to do with it; the farmer follows the price quotations that seem to be made by somebody else, and the buyer of the farm products frequently looks at it in the same way. Although in some cases he has a considerable control over price, frequently the retailer will simply say that "prices have gone up," or "prices have dropped"; he has changed his prices in line with the general market, but he has no idea that he is responsible for this general market change. Individually of course he is not responsible. Yet somehow prices get changed. Through the actions of hundreds and thousands of individual buyers and sellers, prices are pushed up and down, tending always toward a point at which the amount buyers want and the amount sellers offer will be equal. Price is somehow pushed up when buyers want more than sellers offer, and it is somehow forced down when buyers want less than sellers offer. It is this interesting process that is so often glibly described as the "law of supply and demand." Actually the "law of supply and demand" is nothing more than a statement of tendencies toward an equilibrium position in a purely competitive market. When adjustments have been made so that no single individual can, within the available opportunities, improve his position, the situation is stabilized; an equilibrium is attained. Changes will occur so long as this is not the case.

A schedule of the amounts buyers will take at various prices is called a "demand schedule"; analogously, a schedule of the amounts sellers will offer at various prices is called a "supply schedule." A change in demand involves a shift from one demand schedule to another; a change in supply involves a shift from one supply schedule to another. Demand curves generally slope down to the right: more will be taken at low than at high prices. Supply curves generally slope up to the right: less will be offered at low than at high prices.

The concepts of demand and supply are meaningful only in a purely competitive situation. Supply is a set of amounts that will be offered at each of a series of prices, *assuming the seller has no power to raise the*

price. Demand is a set of the amounts that are wanted at each of a series of prices, *assuming the buyer has no power to lower the price.* Purely competitive firms are sellers whose offerings make up supplies; they are also buyers demanding services and products from others. They sell to other firms and to ultimate consumers; they buy from other firms and from owners of "primary productive resources" such as labor and natural resources. A study of the way in which market adjustments would be made in situations approximately purely competitive requires examination of the factors lying back of these various demands and supplies.

Consumer demand has already been studied in some detail in Chapter 3. Examination of purely competitive supplies of the services of primary productive agents such as labor will be deferred until a later part of this book; for the present we merely assume an upward sloping curve with more offered at high than at low prices. In the present chapter, moreover, we shall merely accept the proposition that supplies of goods offered by firms are characterized by larger offerings at higher prices, and that demands by firms are characterized by smaller takings at higher prices. The examination of these propositions will be undertaken in the following chapters. Here we shall merely undertake a brief outline of the processes involved in the functioning of the "law of demand and supply," without answering the important question as to what lies back of the demand or the supply.

Adjustment Toward Competitive Equilibrium

A market is in equilibrium at a price at which the amount offered by sellers in the aggregate equals the amount buyers in the aggregate will take. How does the behavior of the large mass of individual buyers and sellers lead toward this position?

Suppose supply and demand to be as indicated in Table 14—1:

TABLE 14—1

SUPPLY AND DEMAND FOR COMMODITY A

<i>Amount offered</i>	<i>Price</i>	<i>Amount wanted</i>
1,000	\$1.00	4,000
1,500	2.00	3,600
2,100	3.00	3,200
2,800	4.00	2,800
3,740	5.00	2,400
4,675	6.00	2,000
5,610	7.00	1,600

If competing sellers were to start experimentally at a price of \$5.00, what would happen? At \$5.00, 3,740 units will be offered for sale, but only 2,400 units will be taken. Sellers find their stocks moving very slowly, and simultaneously a large number of them "shade" price downward ever so slightly. If only one of them did this he would get all the

trade he wanted and more too, but in fact a lot of sellers are in the same position and they move more or less simultaneously. This general movement brings the market price down. As the price is lowered new buyers appear and buyers already in the market increase their purchases. This process will continue until price drops to a level at which all willing sellers can find buyers for their goods. There will be no reason for any seller to cut price farther.

If, on the other hand, sellers should open with offerings at \$3.00, their goods would sell very rapidly, since at \$3.00, 3,200 units are wanted and only 2,100 are offered for sale. Large numbers of buyers unable to get goods might bid against each other pushing prices up; or sellers might see that they can get higher prices. With the higher prices sellers' offerings would increase and buyers would take less, until a price is reached at which there is no further incentive to change. When this price is reached (1) there are no buyers who would like to purchase any more than they are now buying at this price, \$4.00; no one will offer to pay more for additional units; (2) there are no sellers who would like to sell any more than they are now selling at this price, \$4.00; no one will offer goods for less in order to increase his sales. The market is in equilibrium at a price of \$4.00. At this price the amount offered by sellers in the aggregate equals the amount wanted by buyers.

This situation is presented graphically in Figure 14—1. Demand is

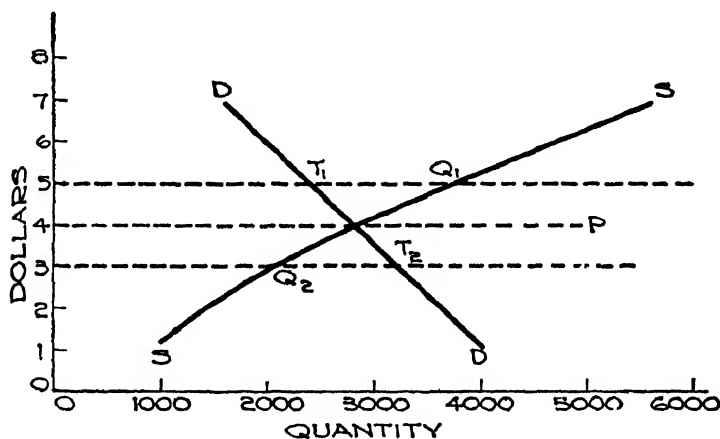


Fig. 14—1. Equilibrium between demand and supply.

represented by the curve DD , supply by the curve SS . At a price of \$4.00 the market is in equilibrium. At any higher price (say \$5) the amount offered, Q_1 , is greater than the amount buyers will take, T_1 ; sellers will cut prices in order to sell their goods. At any lower price, say \$3.00, the amount wanted, T_2 , is greater than the amount offered, Q_2 , and the price will be bid up. The price at which the amount offered and the amount wanted are equal is the equilibrium price.

Effects of Changes in Demand or Supply on Equilibrium Price and Quantity

In the light of this discussion of the competitive pricing process, the effects of a change in demand or supply on price and quantity sold can be readily seen. The simplest way of showing these effects is by the use of graphic representation.

Changes in demand

Suppose the initial supply and demand to be as represented in Figure 14—1. These curves have been duplicated in Figures 14—2 and 14—3. Figure 14—2 also shows an increased demand represented by D_2D_2 . It

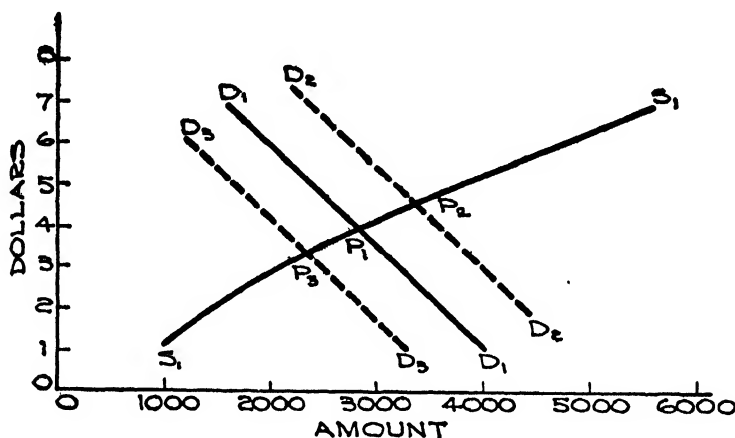


Fig. 14—2. Effect of change in demand.

is greater than the initial demand D_1D_1 , because at each price more will be taken than before. The new equilibrium position will be at a higher price, and a greater amount will be exchanged than before. It takes a higher price to attract the additional offerings of sellers, and this higher price will be offered by buyers who would otherwise (at the price P_1) have been unsuccessful in obtaining the desired goods. Conversely the demand D_3D_3 is less than the initial demand D_1D_1 , and in this case equilibrium price will be lower than before and the amount sold will be less. Sellers finding that they cannot sell all they wish to at the initial price will undercut until the new equilibrium position is reached.

Changes in supply

In Figure 14—3, the supply curve S_2S_2 represents an increase over the initial supply S_1S_1 (more will be offered at each price than before); the associated new equilibrium price will be lower than before and the quantity exchanged will be greater; the supply curve S_3S_3 , which represents a smaller supply than the initial curve S_1S_1 , will lead to a new

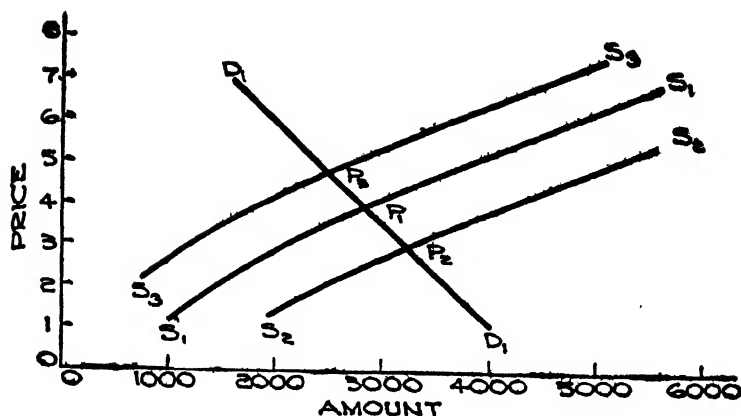


Fig. 14—3. Effect of change in supply.

equilibrium price higher than before, and a smaller quantity will be exchanged than was exchanged when supply was S_1S_1 . The reasons are again those that have been described as determining equilibrium.

Changes in both demand and supply

If both supply and demand change there will of course be a new equilibrium associated with the new seller and buyer positions. Effects

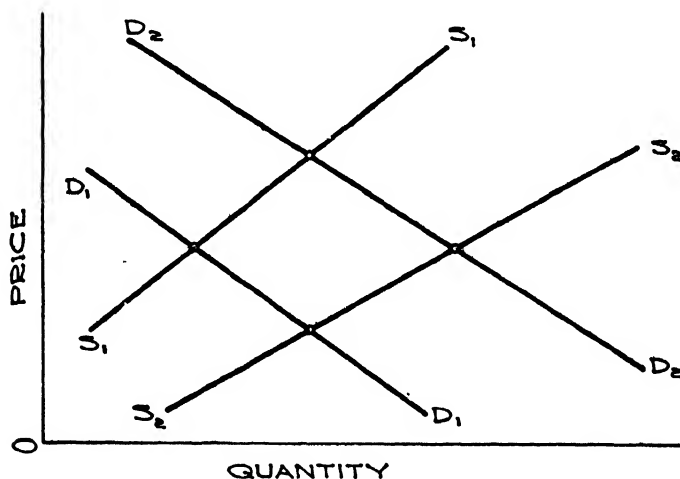


Fig. 14—4. Simultaneous changes in demand and supply.

of such changes on price may work in the same or in opposite directions; effects on quantity exchanged may similarly be mutually supporting or neutralizing. If both demand and supply increase no doubt the quantity exchanged will increase, but price may go either up or down, depending on whether the change in demand or the change in supply has had the greater price effect. If both demand and supply decrease the results will be diminished quantity exchanged, and again an indetermi-

nate effect on price. If demand increases and supply decreases, price will certainly rise, but the quantity exchanged may either increase or diminish depending on whether the change in demand or the change in supply had the greater effect. If the demand decreases and supply increases, price will certainly fall, the quantity exchanged may again either increase or diminish. Some of these variations are illustrated in Figure 14—4; in this figure conflicting price effects exactly balance out, and so also do conflicting quantity effects.

Price and quantity variability

Responses to a change in demand or to a change in supply may be primarily in price or primarily in quantity. If demand is highly elastic, this means that consumers respond readily to price changes by dropping out of the market when prices are raised a little and by coming in and increasing purchases when prices are lowered a little. As a result, most of the adjustment to changes in supply will be adjustments in quantity taken if demand is highly elastic. If demand is inelastic the adjustments will take place primarily in price. Similarly, if sellers respond readily by greatly increasing their offerings with slight increases in price, or by heavy withdrawals with slight price drops, the adjustment to demand changes will be largely in the quantity exchanged. If sellers are quite unresponsive to price in their offerings (if supply is very inelastic), then adjustments to demand change will take place largely through shifts in price.

The Difficulties of Price Fixing

The tendency toward an equilibrium price which has just been described is one of the simplest and most fundamental of economic processes. Efforts to fix prices above or below the equilibrium level lead to many difficulties, which have been summed up by some fatalists in the words "you cannot repeal the law of supply and demand." It is, however, quite possible to change the price if you do something to the supply or the demand or if you intervene with elaborate administrative regulations in the markets.

Attempts to fix prices above an equilibrium level have not been uncommon; minimum wage legislation is under some circumstances such an attempt, and the entire program designed to raise prices on commodities, such as wheat and cotton in the United States and coffee in Brazil, are also examples. Let us consider what would happen if the government stepped in under pressure from wheat farmers and issued a decree that wheat should not sell below \$1.25 a bushel, while the equilibrium price in the market was \$1.00. The demand for and supply of wheat might be as indicated in Figure 14—5.

When the government sets the price above the equilibrium level, consumers curtail their consumption of wheat from 200 to 180 million bushels and at the same time farmers are encouraged to increase their production of wheat to 230 million bushels since the government has promised them

higher prices. There are only two direct ways in which this high price can be maintained: (1) The government can buy large quantities of wheat to make up the difference between farmers' offerings and consumers' purchases; or (2) it can step in with some kind of control on the actions

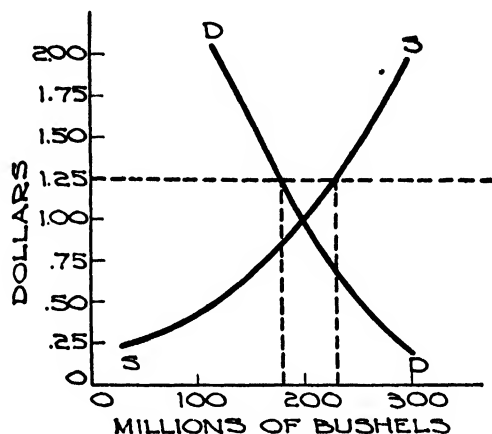


Fig. 14—5. Fixing of a price above equilibrium.

of farmers in producing wheat, thus curtailing their outputs. If the first procedure is followed, the government will build up ever-increasing wheat stocks and will not know what to do with them; this is just what happened to the Federal Farm Board in 1928. The second procedure is what was undertaken under the Agricultural Adjustment program of the 'thirties. Unless the government undertook to support the high price in one or the other of these general ways, the high price would prove unenforceable. Farmers unable to find purchasers for their wheat at the high price set in the government decree would find ways of getting around that price, undercutting until the real price was down to \$1.00 once more.

On some occasions there is an effort to set maximum prices below the equilibrium level, that is, to set a low "ceiling" on the price of a good. This is what the government might do for example to the price of bread during a war period. Assuming the demand for bread and the supply of bread to be as indicated in Figure 14—6, the equilibrium price would be \$.12 a loaf with 6 million loaves sold. But suppose the government were to set a price of \$.10. If the government were to stop at merely decreeing this low price for bread the result would tend to be a curtailing in the offering of bread (back to 5 million loaves if the decree were enforced), and this would leave a large body of unsatisfied consumers. Since there would not be as much bread at \$.10 as the 6½ million loaves consumers wanted to buy, a serious problem would develop in deciding who should get the bread. It might be that long queues would form outside the stores early in the morning on the days when bread came in, and those first in line would get the bread. Perhaps retailers would set a little aside

for especially favored customers. In any case the problem of dividing the bread among the consumers would be a very awkward one and the results would probably be very inequitable. As the government became aware of these difficulties, its program of holding down the price of bread might be made to include either or both of two associated features: (1)

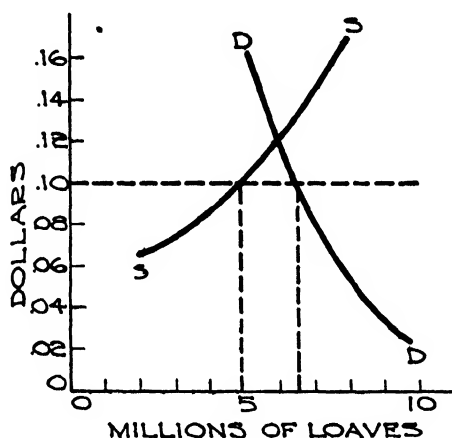


Fig. 14—6. Fixing of a price below equilibrium.

Consumers might be given ration cards permitting them to buy only so much bread each; (2) millers of flour might be paid a subsidy by the government to neutralize the effects of the low prices and encourage the production of large quantities of bread. In this way the government could substitute consumers' ration cards for the rationing mechanism of a free price system; and it could substitute subsidies paid to millers for the price incentives of a free market. If only ration cards were used the quantity of bread offered for sale would be cut to about 5,000,000 loaves. If the subsidy were great enough to bring in enough bread (6,500,000 loaves) to fulfill consumer demand at the low price of \$.10 a loaf there would be no need for the ration cards. In fact the government might use both techniques of control; it could thus check somewhat the withdrawal of bread from the markets by a moderate subsidy, and divide the low-priced bread among consumers through a rationing system.

Periods of Supply

The most important factors determining supply are the actual and anticipated production costs of the sellers. Which costs will be taken into account by sellers will depend on the time period for which they are making decisions. A grocer deciding what to do about offerings of fresh strawberries on Saturday night will not pay much attention to what he had to pay for them, though he would not have purchased them if he had not expected to cover these costs. If a farmer is debating whether or not to pick strawberries he will consider only the added costs of pick-

ing; but if he is deciding whether or not to grow them for next season he will plant them only if he anticipates a price which will cover all the costs involved in doing so. If a manufacturer is considering whether or not to start in business (or expand his plant) he will do so only if he expects that prices in subsequent years will cover *all* the costs involved, but if he is considering what his best policy in the short run will be he may continue to operate so long as he covers variable costs. The lowest price necessary to induce a seller to offer a given quantity for sale will thus depend in part on the time period involved. If the time period under consideration is very long relative to the production processes in an industry, planned sales (outputs) will be based on a consideration of all costs and, looked at in the aggregate, the amount to be offered will reflect the entry and withdrawal of firms in the industry as well as changes in the quantities offered by given firms. If the period planned for is extremely short, only shifts in the outputs of existing firms are possible, and only costs which are variable in the short run will be taken into account.

A schedule of the amounts which sellers would plan to sell in the long run at a series of prices is a *long-run supply schedule*. A schedule of the amounts which they would sell or plan to sell at a series of prices in the short run constitutes a *short-run supply schedule*. Any discussion of factors determining the quantities and price of goods must take into consideration the differences in seller responses with regard to short-run and to long-run plans.

CHAPTER 15

The Firm and Short-run Market Adjustments

ANY single firm in a purely competitive situation adjusts its input-output policies to given market prices as described in Chapters 10, 12, and 13; but, as we indicated in Chapter 14, it is nevertheless the actions of all the many individual competitive buyers and sellers that determine what these market prices will be. The factors lying back of purely competitive consumer demand have already been considered. The present chapter considers the short-run determination of purely competitive supplies of products offered for sale by firms, and the short-run determination of purely competitive demands for the productive services purchased by firms.

The Firm and Purely Competitive Supply

Aggregate market supply is the sum of the supplies of the individual firms that make up the market aggregate. We shall turn first, therefore, to the individual firm; following that analysis we shall look at the aggregate picture.

The supply schedule of an individual firm

The amounts that a particular firm will offer for sale in the short run at each of a series of prices for its product will depend on cost conditions in that firm. We shall assume the cost conditions of the hypothetical hose-manufacturing firm discussed in preceding chapters. In determining the short-run supply schedule for this particular firm we then must ask two questions with regard to adjustments at each possible price for the product: (1) What is the lowest selling price at which the firm will continue to operate in the short run? (2) What will be its most profitable (or least unprofitable) input-output adjustment at each price above this level? When we have answered these two questions we shall know how much this firm would offer for sale in the short run at each price for its product.

The quickest way of answering these questions is to use the data which express costs in relation to outputs. We have therefore reproduced in Table 15—1 these selected cost data for our old friend the hose producer.

TABLE 15—1

SELECTED COST DATA FOR A HYPOTHETICAL MANUFACTURER OF HOSE

<i>Input</i>	<i>Output</i>	<i>Variable unit cost</i>	<i>Marginal unit cost</i>
1	100	\$1.00	\$. . .
2	350	.57	.40
3	702	.43	.28
4	1,152	.35	.22
5	1,700	.29	.18
6	2,190	.274	.20
7	2,604	.269	.24
8	2,908	.275	.33
9	3,114	.29	.49
10	3,240	.31	.79
11	3,300	.33	1.66
12	3,350	.36	2.00
13	3,395	.38	2.22
14	3,435	.41	2.50

1. A glance at the figures for variable unit costs immediately reveals the fact that at any price below \$.269 this purely competitive firm would shut down. It would not pay to operate even in the short run at a price lower than this, since variable costs could not be covered. Nothing would be offered for sale by an enterprise with these costs unless price were at least \$.269 per pair of hose.

2. At any price above this level the most profitable output may be determined by comparing marginal unit cost with selling price. Since the hose are being sold in a purely competitive market this firm does not have any power to change price all by itself, and when it expands or contracts sales by one pair of hose it simply adds or subtracts the price of a pair of hose from its total revenue. Marginal unit revenue equals selling price. Output will be expanded therefore so long as marginal unit cost is less than price, but not beyond that point. When price is \$.30 this hose manufacturer would minimize his losses by producing 2,604 pairs of hose (employing 7 units of input). He would not employ the eighth unit of input because this would increase his costs by .33 per pair of hose added to output while his revenue would be increased by only .30 per pair of hose added. At a price of .33 it would be a matter of indifference to him whether he produced 2,604 or 2,908 pairs of hose; in order to simplify our problem we shall assume that he always selects the larger output if profits (or losses) are the same at two levels. At \$.48 or more he would employ the ninth unit of input, at \$.80 or more the tenth, at \$1.66 or more the eleventh, and so on.

We are now ready to use these comparisons to set up a schedule that will tell us how much this hose manufacturer would offer for sale at each selling price over a considerable range. Such a schedule is presented in Table 15—2. There are big jumps in this schedule owing to the fact that

TABLE 15-2

INDIVIDUAL SUPPLY SCHEDULE FOR PRODUCT OF A HYPOTHETICAL FIRM

<i>Price</i>	<i>Amount offered</i>
\$.26 and under	Zero
.27 to .32	2,604
.33 to .48	2,908
.49 to .78	3,114
.79 to 1.66	3,240
1.67 to 1.99	3,300
2.00 to 2.21	3,350
2.22 to 2.49	3,395
2.50	3,435

outputs can be changed only in jumps, since units of input cannot be broken up into small fractions. Frequently, in a concrete situation the jumps will not be as striking as this because much smoother adjustments will be possible than in our example; but nevertheless there will be some stepping-stone adjustment such as is indicated here. These data are represented graphically in Figure 15-1.

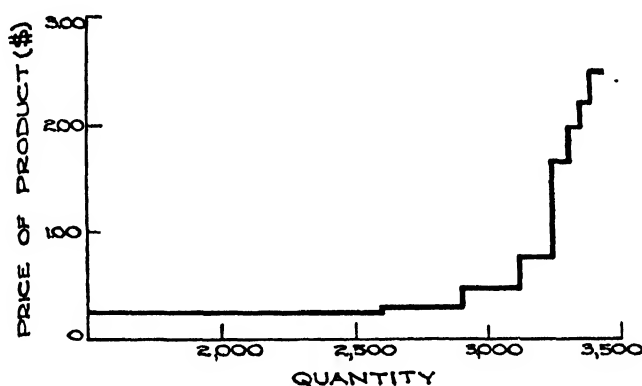


Fig. 15-1. Supply schedule of Firm A.

Aggregate short-run supply

The aggregate supply is the sum of all the individual supply schedules of the separate firms such as the one we examined in the preceding pages. There is no reason to assume each firm's supply schedule to be identical with the others. In fact the existence of many firms which expand or contract their input-output adjustments at different price points results in an aggregate supply schedule much smoother than that for the individual firm. Suppose the firm of the preceding examples to be regarded as Firm A; and that Firms B and C have supply schedules as indicated in Table 15-3. The aggregate offerings of Firms A, B, and C would then be as indicated in the total column of Table 15-3. In the competitive market there will be many more firms than this and the aggregate supply

TABLE 15—3
JPLY SCHEDULES OF HYPOTHETICAL FIRMS A, B, AND C

<i>P</i>	<i>Output A</i>	<i>Output B</i>	<i>Output C</i>	<i>Aggregate A + B + C</i>
.25	Zero	Zero	Zero	Zero
.30	2,604	Zero	Zero	2,604
.40	2,908	1,000	Zero	3,908
.50	3,114	1,000	1,200	5,314
.60	3,114	1,612	1,200	5,926
.70	3,114	1,612	1,915	6,641
.80	3,240	1,612	1,915	6,767
.90	3,240	1,612	1,915	6,767
1.00	3,240	2,104	1,915	7,259
1.10	3,240	2,104	2,532	7,875
1.20	3,240	2,104	2,532	7,875
1.30	3,240	2,500	2,532	8,272
1.40	3,240	2,500	2,532	8,272
1.50	3,240	2,500	2,975	8,715
1.60	3,240	2,815	2,975	9,030
1.70	3,300	2,815	2,975	9,090
1.80	3,300	2,815	3,250	9,365
1.90	3,300	3,050	3,250	9,600
2.00	3,350	3,050	3,407	9,807
2.10	3,350	3,198	3,407	9,955
2.20	3,350	3,198	3,503	10,051
2.30	3,395	3,198	3,573	10,166
2.40	3,395	3,282	3,573	10,250
2.50	3,435	3,282	3,615	10,332

schedule will therefore look much like the relatively smooth curve of Figure 15—2.

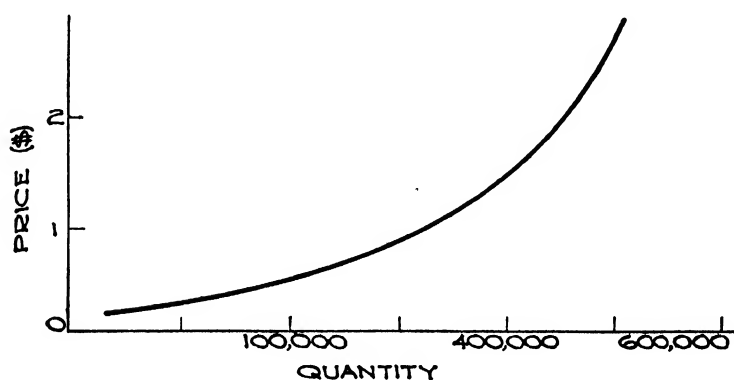


Fig. 15—2. Hypothetical aggregate supply curve.

Preliminary statement of the pricing of the products of purely competitive firms

The price that will actually tend to prevail in the market for the product of these purely competitive firms will depend both on how they would

react at each possible price (supply) and on how buyers respond at different prices (demand). The market price that would tend to prevail will be that at which the amount buyers, in the aggregate, want is equal to the amount sellers, in the aggregate, offer; this adjustment was analyzed in Chapter 14.

The Firm and Purely Competitive Demand for Productive Services

In analyzing supplies we assumed a given set of cost conditions. Analogously, we may assume some given market price for a product, and consider what, under these conditions, would be the demands of purely competitive firms for the productive services that they employ.

The preliminary assumption that a unit of input contains one productive agent only

As a first approach to the problem of the pricing of productive services, a simplifying assumption will be very helpful. We shall assume that a unit of input contains the services of one productive agent only. A unit of input might, for example, be the labor of an unskilled worker for one week. The demand of a firm and the aggregate demand of many firms for units of input is then a demand for that particular productive service. This amounts to assuming that all productive agents but one are fixed. This assumption was made in studying the clock manufacturer of previous chapters; we now apply it to our hose manufacturer as well.

This is a convenient departure from previous assumptions that each unit of input was made up of a "package" of variable agents which was unchanged in composition, in that it enables us to think in terms of the demand for a particular productive service instead of demand for a composite unit or package of productive services. Later we shall take into account the other services that will also be employed by a firm.

The demand schedule of an individual firm

The amounts of a productive service (let us say labor) that a particular firm will take at each of a series of prices for that service (wages) will depend on the revenue opportunities open to that firm. We shall therefore assume the price that can be obtained from the sale of hose to be \$1.80 per pair. The question is now: What quantities of labor will this firm employ at different wage rates? The answer to this question depends on two things: (1) the highest input price (wage level) at which the firm will continue to operate; and (2) its most profitable (or least unprofitable) input-output adjustment at each input price below this level. These are analogous to the two facts we looked for in determining the adjustment of the firm to different selling prices for its product. This time, however, we are taking as given the selling price of \$1.80 per unit of output (average revenue) and considering what the firm will do at various prices of a unit of input (average expenditure).

1. In order to determine the highest input price (wage) at which the firm will continue to operate, we need to find some kind of comparison analogous to that between average revenue, or selling price, and variable unit cost. Average expenditure, or the price of a unit of input, is obviously one of the needed data. If we assume inputs to be labor only, this average expenditure is a wage. Then if there is any adjustment at which the revenue per worker will exceed the wage that must be paid per worker this firm will operate; otherwise it will shut down and no workers will be hired. Putting this in more generalized terms; *if there is any production level at which the revenue per unit of input exceeds the price per unit of input (average expenditure), continuing operation in the short run will pay.* Revenue per unit of input is, of course, total revenue divided by the number of units of input. These figures are given in Table 15—4. The

TABLE 15—4

<i>Input</i>	<i>Total revenue (when average revenue is 1.80)</i>	<i>Revenue per unit of input</i>	<i>Marginal revenue</i>
1	\$ 180	\$180	\$...
2	630	315	450
3	1,264	421	634
4	2,074	515	810
5	3,060	612	986
6	3,942	657	882
7	4,687	655	745
8	5,234	654	547
9	5,605	623	371
10	5,832	583	227
11	5,940	540	108
12	6,030	503	90
13	6,111	470	81
14	6,183	442	72

input, total revenue, and marginal revenue data are taken from previous tables describing the revenue position of a hypothetical purely competitive hose manufacturer when hose are selling at \$1.80 a pair. The column headed "revenue per unit of input" is found by dividing inputs (column 1) into the associated total revenues (column 2).

With a glance at Table 15—4 we may now conclude that the *highest* wage at which this firm would continue to operate in the short run is \$657. At any wage higher than this the revenue obtained per worker would be less than the wages paid per worker; revenue per unit of input would be less than the price per unit of input. Under such circumstances this hose manufacturer would be unable to cover variable cost and would shut down. At any wage rate less than \$657, however, he could get back in revenues something over and above the variable (labor) costs incurred in operating.

2. In answering the question as to what input levels would be most profitable at different wage rates (average expenditures) we are back on

familiar ground once more. One way of answering this question is to compare marginal costs with marginal revenues. Since this firm is buying under purely competitive conditions, marginal costs equal average expenditure; that is, the marginal costs equal the price per unit of input. When the price per unit of input (the wage of the hosiery workers) is \$100, each addition of a unit to input adds \$100 to the cost of the firm; when price per unit of input is \$200, each addition of a unit of input adds \$200 to the cost of the firm; and so on. We may therefore determine the most profitable input by comparing the price per unit of input with marginal revenues. So long as the price of a unit of input is less than the marginal revenue associated with its employment the firm will expand; beyond this point it will not go. Thus if our hosiery workers were to receive wages of \$600, 7 would be employed. The seventh worker adds \$745 to revenue and only \$600 to cost so he will be hired; but the eighth worker adds only \$547

TABLE 15—5

HYPOTHETICAL HOSIERY FIRM'S DEMAND SCHEDULE FOR A PRODUCTIVE SERVICE

<i>Price per unit of input</i>	<i>Input</i>
\$658 and over	Zero
657-548	7
547-372	8
371-228	9
227-109	10
108- 91	11
90- 82	12
81- 73	13
72-	14

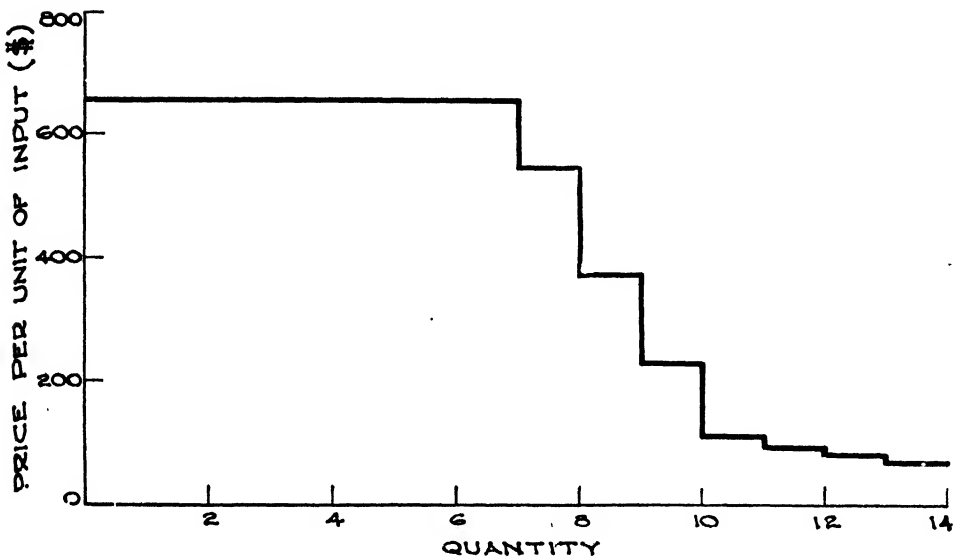


Fig. 15—3. Hypothetical firm's demand for a productive service.

to revenue and would therefore not be "worth" his wage. If the wage rate were \$547, it would be a matter of indifference whether the firm stopped with the employment of the seventh worker or hired 8. If wages were as low as \$200, 10 workers would be hired, and at wages of \$100, 11 would be employed. (All of these conclusions could easily be checked by any disbelievers, simply by figuring out total costs on each wage assumption and comparing the results with total revenues. Such figures are presented for an input price of \$100 in Table 10—1A of Chapter 10, and for an input price of \$600 in Table 10—1C of that chapter.)

We are now in a position to draw up a demand schedule for hosiery workers in our hypothetical enterprise. At any wage level above \$657 this firm would go out of business; employment at wages below this figure is found by comparing the wage (which under pure competition equals marginal cost) with marginal revenue. The results are given in Table 15—5, and presented graphically in Figure 15—3. Again there are, of course, jumps in this schedule due to the indivisibility of a unit of input.

Aggregate short-run demand

The aggregate demand for a productive service is the sum of all the demands of individual firms using that productive service. They may be engaged in the same industry or in different industries. The aggregate demand for cleaning-women in the city of Chicago is obviously made up of the demands of hundreds of different firms engaged in different productive activities. There is again no reason to assume each firm's demand schedule to be identical with the others. The existence of many firms that expand or contract their input-output adjustments at different average expenditure levels results in an aggregate demand schedule much smoother than for the individual firms. An aggregate demand schedule which included our hypothetical firm along with many others might look very much like the curve drawn in Figure 15—4.

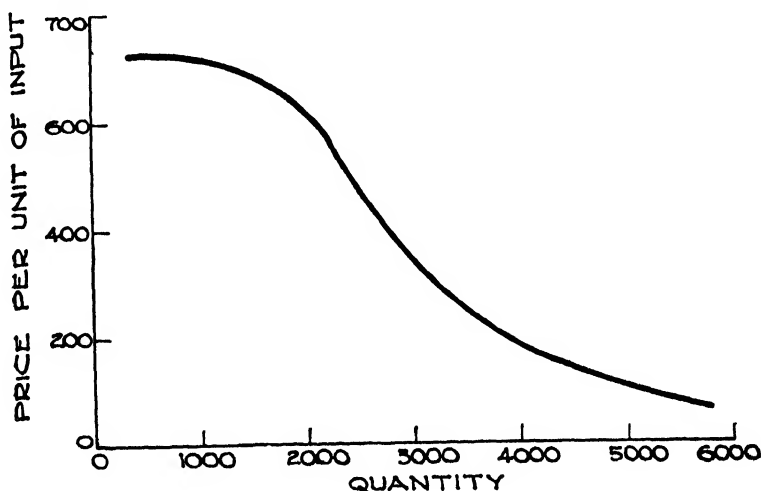


Fig. 15—4. Hypothetical aggregate demand curve.

A Note on Partial Equilibrium Analysis

The analysis of this chapter has been what is called in economics "partial equilibrium analysis." We hold certain factors "constant" and consider what would happen with changes in the particular factor on which we are focusing attention. Thus in constructing supply schedules we assumed given prices per unit of input and hence given cost schedules. In constructing demands for productive services we assumed given prices of the product and hence given revenue schedules. The assumption that a unit of input was made up of one productive service only was merely an aspect of this type of "partial analysis." Actually these things are all closely interrelated. Through a sort of juggling back and forth adjustments tend to be made until in all markets buyers and sellers are so situated that no change available to them would improve their positions. These interdependent adjustments will be considered more fully in subsequent chapters.

CHAPTER 16

Long-run Price and Output Adjustments

THE essential difference between short-run and long-run adjustments is this: Short-run adjustments are merely the adjustments of existing firms in their inputs of variable agents and their outputs of products. *Long-run adjustments in an industry take place through the entry or withdrawal of firms* (and hence of the resources which they employ). They are adjustments which take place over a time period long enough to permit such a shifting of firms, and of relatively fixed productive agents.

For many purposes it is more important to analyze long-run adjustments under purely competitive conditions than to analyze the short-run adjustments in markets characterized by pure competition. Long-run analysis provides a rough clue to the *ultimate*, in contrast with the immediate, effects of introducing some new policy in the economy, of shifts in consumer demand, of improvement in technology, of other lasting changes. If a tax is imposed on housing, for example, we shall want to know not only what would be the effects on rents immediately after imposing the tax, but what would be the ultimate effects after there has been time for the construction industry to curtail house production so that the available amount of housing is diminished. If a tariff is imposed on sugar what will happen to the price and the consumption of sugar, not only this year and next year, but five years from now; and how will this affect the use of labor resources in the South? How will the development of a new and more efficient machine process for the spinning of wool ultimately affect the quantity and prices of woolen goods, and the employment and wages of workers in the woolen goods industry?

The various parts of an economy might be imagined to move toward some sort of equilibrium adjustment as a much harassed traveler in a war-ridden world moves toward his destination. The traveler may start for France, but changes his plans because Paris has fallen to the Germans. He then decides to go to Spain but the boat is driven off its course by a bad storm. This leads him to head for the west coast of Africa but his boat is torpedoed and he is rescued by a Dutch vessel which is headed for Brazil. The chief difference between the economy and this traveler is that on the one hand the traveler's reversals are more drastic than would usually occur in economic changes, and on the other hand the traveler will probably eventually reach at least one of his destinations; whereas the frustrations

in economic adjustments are likely to go on forever. While entrepreneurs producing silk goods are adjusting to an increase in demand for their product a new competing good appears on the market, or the supply of raw silk is curtailed, or a new technique in silk goods manufacture alters their course, or perhaps consumers change their minds about silk goods. Yet at each step along the way a chart of the planned course of action would have been a useful guide in understanding and predicting the direction of movement. With each change of conditions a new chart would be required.

A time period in economics is a concept and not any specific historic lapse of time. It is a period which permits the completion of a part or the whole of an uninterrupted journey from some given starting point to destination: The short run is just one lapse of such a journey, leading to a very temporary equilibrium position (stopping off point). The long run is a complete journey, again carried out without any intervening change of conditions which would alter its course, and the end of the journey is a long-run equilibrium (travel will cease). In studying long-run adjustments we are thus examining what would happen if time were given for firms to shift from one industry to another, and during this interval no new influences were injected into the picture. •

The essence of adjustments toward long-run equilibrium is the movement of resources from less remunerative to more remunerative uses, under the direction of entrepreneurs desirous of making profits. We may therefore define equilibrium of an industry either in terms of the movement of resources or the movement of firms. *An industry is in equilibrium when there is no advantage to any productive agent in moving into or out of it, or (alternatively) when there is no incentive for entrepreneurs to inaugurate or withdraw firms from it.* This equilibrium of an industry with other parts of the economy is a kind of "partial" long-run equilibrium.

The purpose of this chapter is the examination of the processes of long-run adjustment in an industry. We are assuming the industry to be purely competitive in both buying of productive agents and selling of the product. We are also assuming that there are no artificial restrictions on the entry or withdrawal of firms and of the productive resources which they employ. In making this study we shall determine where the industry would arrive at the end of the journey if no new influences had intervened to change its course.

The Process of Adjustment Toward Long-Run Equilibrium

Response to profit expectations

The analysis of the preceding chapters has been largely concerned with the activities of entrepreneurs in the "short run"—in that period of time in which certain costs, usually thought of as connected with fixed investment in plant and equipment, must be taken as fixed and unavoidable, regardless of output policy. As we have seen, a firm may be in its best possible short-run position and be making a profit, suffering a loss, or just

breaking even. Since the price is outside the control of the individual entrepreneur under pure competition, his activities in reaching his most advantageous position consist in adjusting his inputs and outputs to the most profitable relation with the given market price. If in this adjustment he is making a profit, he is fortunate. If he is making a loss, he is unfortunate. But whichever the case may be, there is nothing more he can do about the matter, given his existing fixed costs.

In the "long run," however, equipment wears out and the firm must either replace this equipment or withdraw from business. Long-term contracts leasing land, or hiring managers, or borrowing funds, and so on, expire, and these contracts are considered for renewal or replacement. In the long run "fixed" costs thus become "variable" and an entrepreneur will compare *all* costs with price in making his decisions. Moreover, if there is no restriction on the establishing of new firms in the industry, the possibility of entry of new firms must be taken into account. Thus any given short-run equilibrium of inputs, outputs, and market prices can only persist if there is no tendency for entrepreneurs either to curtail replacement of the relatively "fixed" agents or to expand these agents and to establish new firms. If entrepreneurs anticipate profits, and entry to the industry is free, new firms will be established. If losses are anticipated, existing firms will not replace worn-out agents or renew expired contracts for the relatively "fixed" agents. *And firms will move into or drop out of the industry until expectations of profits and losses have been roughly eliminated—until it is no longer possible for anyone to better his position by moving into or out of the industry in question.* If this point were reached, we would say that the industry was in long-run equilibrium.

Looking ahead at costs and revenues

Adjustments in production are based on present costs and prices and on anticipations of future costs and prices. The costs significant for planning are those that *will be* incurred if certain plans are followed. This is true whether we are considering short- or long-run adjustments. Of what significance is this fact in the long run? This question may best be answered by the use of an illustration.

Suppose that a machine used in the overcoat industry cost \$5,000 in 1940, and that it was purchased for this sum by a number of firms in the industry. In 1943 machines like this can be purchased for \$4,000. It may be that the firms which purchased the machine for \$5,000 in 1940 are still using it and that such a machine will continue to be useful for another two years, making it now a fixed agent. Firms that purchased the machine for \$5,000 may be unable to cover their total costs, including the entire \$5,000, and typical existing firms with these costs may be operating at a loss. We may now ask two questions: (1) Will existing firms replace this machine when it wears out (say in 1945)? (2) Will new firms enter the industry even though old firms having purchased the machine in 1940 are operating at a loss? The answer is this: if entrepreneurs considering either reinvestment or entry to the industry anticipate that on the basis

of present costs they will be able to make a profit (or just cover costs) they will decide to make the investment. The 1940 cost of \$5,000 for the machine is past history, and does not enter into present estimates of profit possibilities. The fact that firms which incurred this cost of \$5,000 in the past were unable to cover it does not mean that purchasing the machine now for \$4,000 would not be profitable. Thus it is possible that in an industry where existing firms are "losing money" it may still be profitable for old firms to replace investments in fixed agents when they wear out and for new firms to enter the industry.

Conversely, if the price of the machine has risen it might be that old firms making profits on the basis of the costs they paid in the past would not be able to make profits if they reinvested in replacing worn out equipment. In such circumstances there might be some withdrawal of old firms even though existing firms were showing profits over previously incurred costs.

To restate the conclusions of this section: the costs which will be taken into account in estimating profit opportunities are the present and anticipated costs of the productive agents to be employed. These are the costs which are anticipated by any new firm or by any old firm replacing its equipment.

Illustration of response to an increase in demand

In order to understand the exact nature of long-run adjustments we may begin with a preliminary example of an industry that is in an equilibrium position; there are no unexploited profit opportunities for new firms in this industry, and there are no old firms that would withdraw as their plant and equipment wore out. Then let us assume that demand for the product of this industry increases. We shall suppose that this is a purely competitive industry producing desk blotters.

The immediate effects of the increase in demand for desk blotters will be an improvement in the opportunities of the firms already in the industry. These short-run adjustments are easily traced: (1) The price of blotters will be higher than before, since nothing has happened that would cause the firms in the industry to offer them more readily at any given price than previously, while the buyers have an increased demand. (2) Each firm will tend to increase its output, since with a higher price it is now profitable to add to output units that previously would have added more to costs than to revenue. (3) As a result of the higher price and greater output, the profits of each of the blotter firms will be increased.

This short-run adjustment is easily illustrated by the use of graphic analysis. Figure 16—1A gives the hypothetical short-run aggregate supply of desk blotters, the initial demand (D_1D_1) and the new increased demand (D_2D_2). The increase in demand has the effect of calling forth an increased output from the firms already in the industry, 60,000 instead of 50,000 blotters, and of raising price from \$.06 to \$.08. This is a result of the simultaneous actions of the many purely competitive firms producing blotters. To the individual firm the increase in demand makes

possible a higher price, which calls forth from each firm an increased output of blotters; marginal unit cost equals marginal unit revenue (price) at a larger output than before. The situation of a typical individual firm is illustrated in Figure 16—1B. The output in this firm rises from 400 to 500 blotters. The aggregate of such increases by all firms in the industry is, of course, the increase by the industry from 50,000 to 60,000 blotters. And profits are now obtained, as shown by the shaded area in Figure 16—1B.

Thus far we have been reviewing the short-run adjustment that sets the stage for those changes that are spoken of as "long-run" adjustments. Profit opportunities in the industry have appeared. What will happen

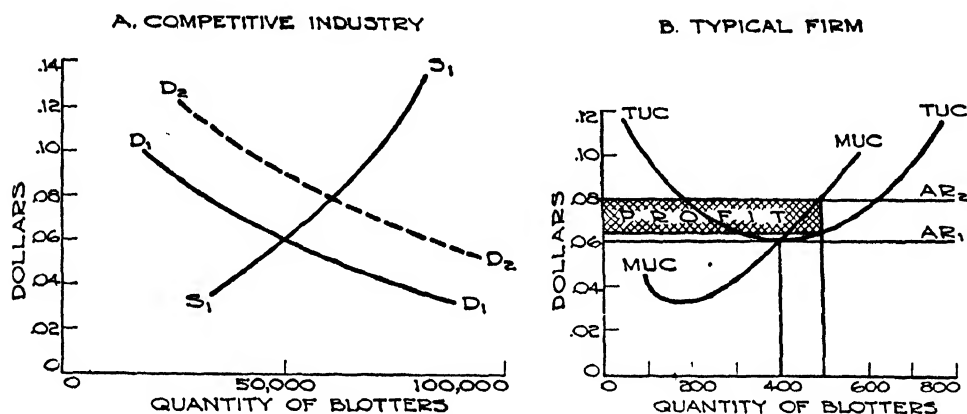


Fig. 16—1. Adjustment prior to entry of new firms.

when sufficient time has elapsed (no other changes intervening) so that new firms may enter the industry?

As businessmen who are attracted by the possibilities of making profits in the production and sale of blotters secure productive agents and organize new firms, the aggregate output of blotters will increase. There will be a productive capacity in the industry prepared to produce a somewhat larger output over a whole range of prices. The new short-run aggregate supply (once the new firms are established) will have increased; in terms of graphic analysis, the new short-run aggregate supply curve will have moved to the right.

So long as businessmen can enter the industry and sell blotters at a price above their lowest total unit costs, there will continue to be an inducement for new firms to enter. *But as more firms appear two forces are brought into operation that tend to reduce and eventually to eliminate the profits earned by these firms.* The first of these is a reduction in the price of the product, which is a consequence of the increasing supply. The second is that *the level of costs is raised for all firms* as higher and higher payments must be made to attract productive agents. With the price of the product falling and costs rising, profits are gradually "squeezed out," and when the point is reached where the price is equal to the anticipated lowest total

230 LONG-RUN PRICE AND OUTPUT ADJUSTMENTS

unit cost of a new firm, there will be no further inducement for the entry of new firms.

These adjustments are illustrated graphically for both the firm and the industry in Figure 16—2A. The curve S_2S_2 in Figure 16—2A is the new short-run aggregate supply curve after new firms have had time to come into the industry. Under these conditions supply and demand are equal at a price of \$.07 per blotter, with an output of 70,000 blotters in the industry as a whole. This is the adjustment that will tend to appear after long-run forces have taken full effect. The position of the typical firm will now be changed as shown in Figure 16—2B; the cost curves will have risen; and the price, or average revenue, will be greater than before the increase in demand, but less than when the first impact of that demand was felt (that is price will have risen from \$.06 to \$.08 and then fallen back to \$.07). The typical firm will be producing just at its lowest

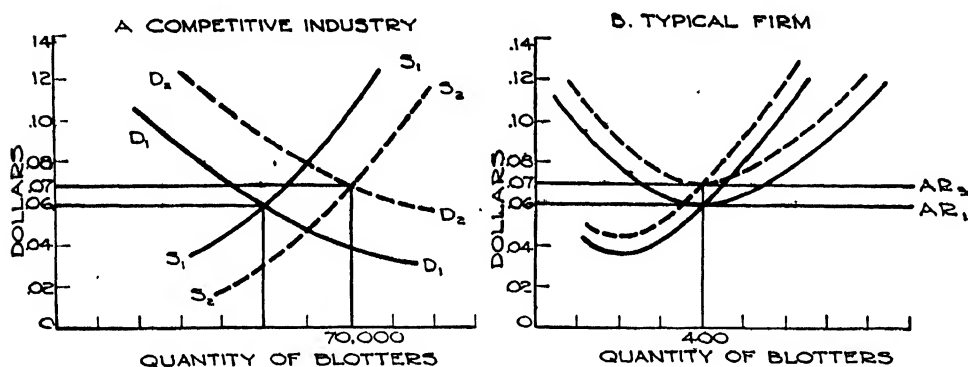


Fig. 16—2. Adjustment after entry of new firms.

total unit cost point (instead of beyond that point); but since there are more firms than before, the aggregate output of the industry is much greater than it was initially.

Thus we see that in the long run there will be a greater expansion in the output of the industry than in the short run, when there is no change in the number of firms. In the short run individual producers in the industry operate their fixed plant and equipment more intensively; in the long run more capacity is added to the industry (through the entry of new firms), and the aggregate output is increased until price drops back to the point where it equals the lowest total unit cost of a new enterprise. This new lowest total unit cost will usually be greater than with the smaller output of the industry (before the increase in demand).

The adjustments which would have followed a decrease in demand would have been the exact reverse of those we have traced through. A decrease in demand from D_2 to D_1 would have had the following results: Price would have fallen at first from \$.07 to \$.05 and then risen back to \$.06 as firms withdrew from the industry; output would have declined to 60,000 and then to 50,000 blotters. In the short run, firms in the industry would have suffered losses; after some of them had been

eliminated those remaining would find their costs less and price modified upward so that they would be able to cover costs once more.

The long-run supply schedule

The type of supply curve with which we have been working hitherto serves to indicate the amount of the product which would be placed on the market at any given price *in the short run*. Graphically we may show the long-run adjustment by a series of shifting short-run supply curves. Supply curve S_1S_1 in Figure 16—2A was the short-run curve that would prevail in the blotter industry if that industry were in long-run equilibrium under the conditions of demand represented by the demand curve

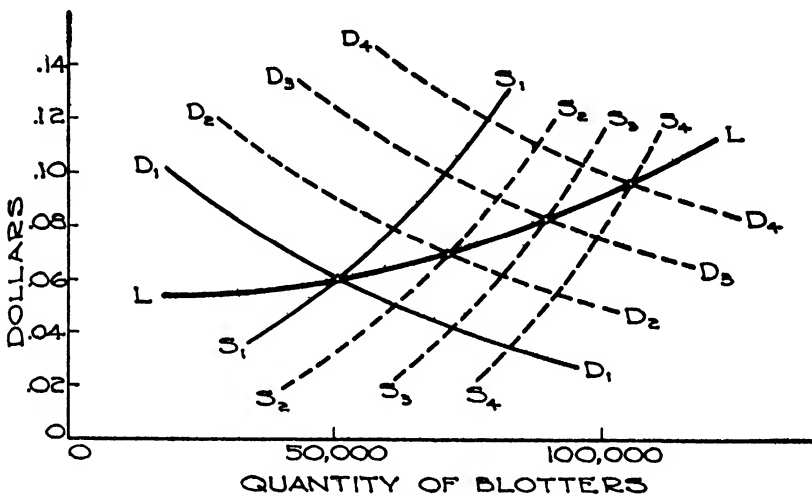


Fig. 16—3. Derivation of long-run aggregate supply ("LL").

D_1D_1 . Supply curve S_2S_2 is the short-run curve that would prevail if the blotter industry were in a long-run equilibrium position under the demand conditions represented by demand curve D_2D_2 . We might have drawn an entire series of short-run supply curves that would accompany the long-run equilibrium adjustments under various demand conditions. This would be a very awkward procedure, however, and our graph would be cluttered up with a large number of curves.

There is a much simpler possibility. We can construct a single long-run supply curve each point on which will show the price at which any given output of the industry would be forthcoming *after all businessmen able to secure productive resources and produce at that price without loss have become a part of the industry, and after all others have dropped out*. This is simply a curve drawn through the points of intersection of the appropriate short-run supply curves and demand curves. This is shown in Figure 16—3. The curves $S_1S_1—D_1D_1$ and $S_2S_2—D_2D_2$ are taken from Figure 16—2A; others are added. The curve LL is the associated long-run supply curve for this industry.

Relation Between Costs and Long-Run Supply

Survival of the fittest and long-run uniformity of lowest total unit costs

At any given time there may be firms of varying size and efficiency in an industry, some possibly making profits, and others incurring losses. But as long as it is open to anyone to enter this industry freely, an excess of price over attainable minimum total unit costs will encourage the entry of new firms. As such firms move in they compete with existing firms, and the most inefficient are eliminated. In the long run there will remain in the industry only firms that have lowest total unit costs as low as those that would be incurred by new enterprises in optimal scale adjustments. *If a long-run equilibrium position were finally attained, there might still be many differences between firms, but the lowest total unit costs of all firms would be the same.*

While it is easy to see that in the long run inefficient firms will tend to be eliminated by competition, it is contrary to observation to assume that all firms tend to become identical. Some entrepreneurs are more efficient than others; some firms are located near good markets and pay high rents while others are more distant but pay lower rents; some firms are small, with close personal supervision to obtain efficiency, while others are large, with mass production methods counted on to lower total unit costs. Such manifold differences may of course exist in long-run equilibrium. It is in no way necessary that all firms be identical or have identical cost curves. It is only necessary that the methods of production used by each permit it to produce at a total unit cost as low as its competitors. *The persistence of differences between firms along with the tendency toward uniformity of minimum total unit costs is due to the balancing of compensating advantages and disadvantages.* This is easily illustrated.

Suppose two manufacturers of cotton textiles to be differently located. One may have the advantages of location near buyers but in that case entrepreneurs competing for the use of such land will bid rents up to a relatively high level. The other may be located where transportation costs to market are higher, but the rents it pays will then tend to be lower. In the long run the result of competition for the use of land will be that difference in the rent or sale price of the land.

Or suppose that one manager of a textile-manufacturing firm is far more effective in handling men and organizing production than is any other person in the industry, so that his firm makes substantial savings not available to other producers. Will not this particular firm continue to make a profit even in the long run, its costs remaining below the costs of other firms, since none of the others can duplicate its efficiency? The answer is, No. The firm operated by the more efficient manager must count as a cost a higher salary, which neutralizes the savings provided by his superior services. If the manager is hired, the firm will have to pay him a higher salary than other managers receive in order to bid him away

and keep him away from other firms; his salary will be bid up until it is higher than that of a less efficient manager by just the differential advantage of his services. If the entrepreneur himself is the efficient manager, he must charge as a cost a salary for himself equal to what he would be able to get in alternative opportunities; here "implicit cost" would be the compensating factor.

Thus in the long run, competition for productive resources will tend to force an adjustment that balances the differences in productivity between firms. All firms in an industry will tend to have the same lowest total unit cost. Therefore in speaking simply of the lowest total unit cost of an optimum scale firm, we are describing all the firms in a long-run adjustment.

The long-run aggregate supply curve—a long-run cost curve of the industry

In describing adjustments in the blotter industry, we depicted a typical firm as just covering costs when the industry was in a position of long-run equilibrium. It is now evident that this firm represented all the firms in the industry insofar as its position in the final adjustment was concerned. If a position of long-run equilibrium were attained no firm could make a profit and none would incur a loss. If profits were possible new firms would enter until profits were eliminated, and vice versa; and since all firms would in the end have the same lowest total unit costs all would be "in the same boat." (It should be remembered that this does not mean that entrepreneurs and investors would be starving. They would simply be receiving incomes on their services and investments just equal to what they could get in alternative occupations; this might mean very substantial incomes even in the absence of profits in the "economic" sense.)

Since in a position of long-run equilibrium each firm would be just able to cover costs, each would also be producing at optimum scale and at its least-cost-combination output. At any other output costs would exceed revenues. Incidentally, this level of output is also the output at which marginal unit cost equals marginal unit revenue; it is the special case in which marginal unit cost equals total unit cost. The situation is illustrated graphically in Figure 16-4.

A schedule might be drawn up showing what the lowest total unit cost of a typical firm (and hence of all firms) would be when, with all firms producing at their least-cost-combination outputs, the aggregate output of the entire industry was 200,000 blotters.

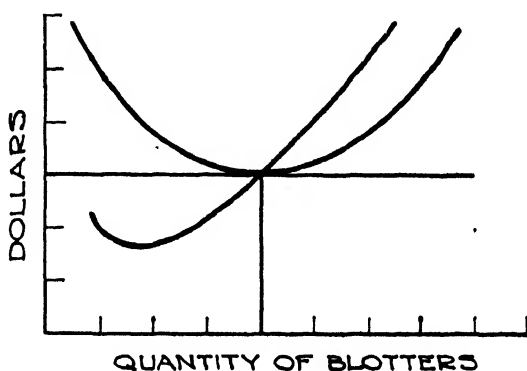


Fig. 16-4. Output adjustment of a purely competitive firm just covering costs.

234 LONG-RUN PRICE AND OUTPUT ADJUSTMENTS

or 300,000 blotters, or 400,000 blotters. This is called a *long-run cost schedule of the industry*. It is also a statement of the prices that would be necessary to call forth these various outputs of the industry in the long run, hence it is a long-run aggregate supply schedule. A curve representing long-run supply in the hypothetical example of the blotter industry has already been presented. Such a curve might be based on the cost data of Table 16—1. Each point on the long-run aggregate supply curve

TABLE 16-1

LONG-RUN AGGREGATE SUPPLY SCHEDULE	
<i>Output of the industry (all firms operating at lowest TUC)</i>	<i>Lowest TUC of an optimum size firm</i>
.....	\$..
200,000	.12
300,000	.13
400,000	.15
500,000	.18
600,000	.22
700,000	.27

thus tells us the lowest total unit cost of a typical firm at that aggregate level of production, assuming each firm to be contributing just its least-cost-combination output.

A crude graphic representation of this relation between cost and long-run supply of the industry is given in Figure 16—5. The horizontal axis

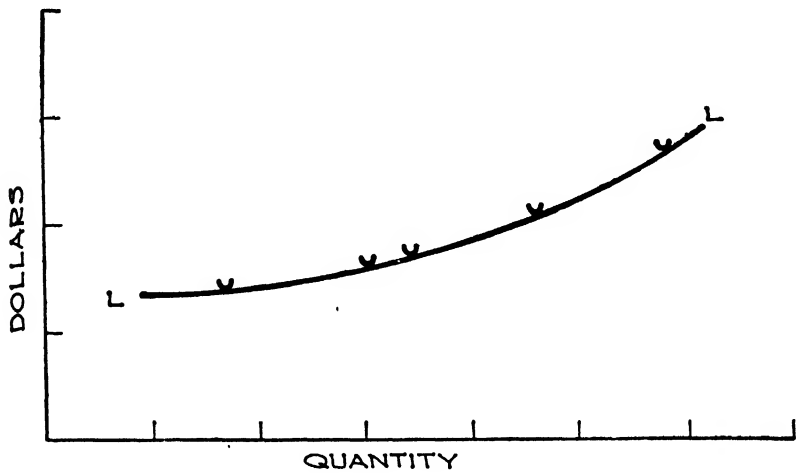


Fig. 16—5. Long-run aggregate supply; cost in typical firms.

is drawn on a very small scale, showing the aggregate output of the entire blotter industry. Since this is assumed to be an industry characterized by pure competition, each firm would be very tiny in relation to the entire industry. At any given output of the industry'as a whole the total unit costs of a single optimum scale firm might be represented by a tiny curve

such as any of the curves drawn at random along the long-run aggregate supply curve of Figure 16—5. Each of these tiny curves will rest at its lowest point exactly on the aggregate curve (*LL*); each tiny curve simply represents all firms when the long-run aggregate output of the industry is at a certain point. Thus when the long-run aggregate output of the industry is 200,000 blotters, the lowest total unit cost of a typical firm is \$.12; when the aggregate output is 300,000 the lowest total unit cost of a typical firm is \$.13; when the aggregate output is 400,000 the lowest total unit cost of a typical firm is \$.15.

Relation between costs to a firm and the size of an industry

In most industries, a general expansion of the industry (including the entry of new firms) will result in higher costs for each firm in the industry; these are called *increasing cost* industries: the larger the aggregate output of the industry as a whole, the higher the costs of production in each and every firm in the industry will be. There are two reasons why an expansion in the output of an entire industry leads to higher costs: (1) The prices of productive services are likely to rise. (2) Additional agents are likely to be less efficient.

1. Any one firm in a purely competitive industry could expand without exerting an appreciable effect on the prices of the productive services it employs; but when many purely competitive firms are appearing and expanding at once the aggregate effect will usually be to bid up the prices of the services of the productive agents which they use. In part this bidding up of the prices of productive services is a bidding of the firms in the industry against each other in their efforts to get a share of the limited quantities of productive services available; in part it is a bidding up of prices of productive services that are attracted away from other industries. Conversely, when there is general contraction in an industry, agents specialized to that industry undercut each other in the effort to retain employment, and the prices that will be paid for their services by any firm in the industry will fall. As a result of these mass effects, of simultaneous expansion or contraction by many firms in an industry, prices paid for productive services by each firm will usually be higher when the aggregate output of the industry is large than when it is small.

2. New agents attracted into expanding industries are frequently less efficient than those initially employed. Perhaps land previously regarded as "sub-marginal" is now cultivated, or low-grade copper mines are opened up. It is true that in any given market these relatively inefficient agents command lower prices than do the better grade agents, but they also produce less.

As a result of these two tendencies, the lowest total unit cost of an optimum scale enterprise is likely to be greater when the output of the industry as a whole is large than when it is small, that is *most industries are increasing cost industries*. It will therefore take a higher price of the product to attract a large number of firms into an industry, and to main-

tain them there than is necessary to maintain a smaller number of firms

There may be exceptions to this situation of increasing costs with expansion of an industry. Some industries are very small, and though they may be made up of a large number of firms they may still employ a very small part of the total of each of the productive services that they use. In such circumstances some expansion or contraction in the industry as a whole may take place without resulting in any appreciable change in costs. Such cases are described as *constant cost* industries: the prices paid for the use of raw materials and labor and land and all the rest of the agents used in the industry remain unchanged as the industry increases its demand for these productive services. This situation is very unlikely, however. Even in extremely small industries some productive agents will probably be bid up in price as firms in the industry seek simultaneously to increase their outputs and as new firms appear and compete with them for these productive services. And as the industry gets larger and larger it must take a more and more important place in the economy, until further expansion will inevitably involve increasing costs.

A Re-Examination of Influences Determining Supplies of Products

Elasticity of supply of a productive service and supply of the product

Firms in the cotton textile industry employ semiskilled workers to spin and weave, skilled mechanics to tend and repair machines, unskilled women to clean. The firms buy raw cotton, and coal, and looms, and spindles, and office supplies. They buy the services of members of the construction industry in the building plant, and they buy the brick and wood and glass and steel that go into their buildings. They buy or rent land. The spinners and the weavers and the raw cotton are relatively specialized to the cotton textile industry; but the mechanics and cleaning-women and coal and builders' labor and office supplies are not. An expansion in the size of the cotton textile industry would have but a minor effect in the markets for the labor of cleaning-women or the buying and selling of coal. An expansion in the size of the cotton textile industry would have a very important effect in the markets for the labor of spinners and weavers and raw cotton. An expansion in the size of the cotton textile industry could take place without exerting any significant effect on the price of coal because textile firms buy such a small proportion of the total purchases of coal; the whole cotton textile industry could increase or decrease its purchases of coal approximately at the price of coal; in technical terms, *the supply of coal to the cotton textile industry is highly elastic*. An expansion in the size of the cotton textile industry would, however, in all probability exert a significant influence on the wages paid to weavers because the purchases of such labor by the cotton textile industry constitute a large proportion of the total amount of labor services of weavers; if the whole cotton textile industry were to expand to

any large extent the wages of weavers would be bid up; *the supply of weavers to the cotton textile industry is much less elastic than is the supply of coal to that industry.* It is because of the increasing payments to such agents that the costs of firms in an industry rise as the industry as a whole expands.

From these rough comments we may now draw two important conclusions: (1) *If a productive service is very inelastic in supply to an industry, its price will rise (be bid up) sharply with expansion in the industry, thus exerting an upward pressure on production costs.* (2) *If a large proportion of the productive agents used in an industry are highly inelastic in supply to that industry costs of production in the industry will rise sharply as the industry expands.* A steeply rising cost curve will mean that large expansions of output of a product will occur only in response to very great increases in the price of the product; in other words, *when the major resources used in an industry are highly inelastic in supply to that industry the derived supply of the product will be inelastic.*

Exactly the opposite results will follow if the supplies of productive agents to an industry are very elastic. In fact probably no productive resources used in the cotton textile industry are extremely inelastic in supply to the industry. The long-run cost curve of the industry rises only slowly; the aggregate long-run supply of cotton textiles is very elastic. *When the major resources used in an industry are highly elastic in supply to that industry the derived supply of the product will be elastic.*

In using the realistic example of the "cotton textile industry" we have done violence to precise use of economic concepts (this always happens in concrete applications); but our conclusion is none the less valid. The "supply of cotton textiles" is a somewhat nebulous thing; the supply of cotton sheeting of a given size and quality would be more precise in meaning. If we hypothecate a purely competitive industry producing cotton sheets all of the same size and quality we may apply our analysis more accurately. If the hypothetical supply schedules for offerings of all productive agents to the cotton sheeting industry were highly elastic, the aggregate supply of cotton sheeting would be also; if these supplies of productive agents to the industry were highly inelastic, the aggregate supply of cotton sheeting would be inelastic as well.

Influences determining the supply of a productive service to an industry

The total supply of raw cotton is offered to buyers who want to use it in many different ways; in making dresses, and shirts, and draperies, and rugs, and rag dolls. How much cotton will be available at various prices to firms making rag dolls will depend both on what happens to the total supply of raw cotton for all uses and to the demand for cotton in uses other than the making of rag dolls. *The greater the aggregate supply of raw cotton the greater the supply of cotton to the rag-doll industry; and the greater the demand for raw cotton in other uses the less the supply to the rag-doll industry.* The more specialized a productive agent

is to an industry the more directly does its supply to the industry reflect the total supply of the agent.

Similarly, the elasticity of supply of raw cotton to the rag-doll industry will depend on both the elasticity of the total supply of raw cotton and the proportion of cotton that is habitually used by the rag-doll manufacturers. *The more elastic the total supply of raw cotton, the more elastic the supply to the rag-doll industry; and the greater the proportion of cotton habitually consumed by the rag-doll industry the less elastic the supply of cotton to the industry.* In the particular case of the supply of cotton to the rag-doll industry, we should expect a high degree of elasticity because the industry takes such a small part of the total amount of cotton generally purchased; this is more important in this case than the relatively inelastic character of the total supply of cotton to all uses. In contrast, spindles adapted to spinning coarse cotton thread would be elastic in supply to manufacturers of coarse cotton textiles because the total supply of such spindles is very elastic; this is more important in the case of the spindles than is the concentration of purchase of the spindles in the single industry (coarse cotton goods).

Summary of the factors influencing the supply of a product

The conclusions of the last two sections may now be drawn together in a brief summary:

1. An increase (or decrease) in the total supply of a productive agent, especially when it occupies an important place in a particular industry, will tend to increase (or decrease) the supply of the product; production costs in each firm will be lowered (or raised).

2. A decrease (or increase) in the demands for a productive agent in other uses will tend to increase (or decrease) the supply of the agent to the particular industry under consideration. This will tend to increase (or decrease) the derived supply of the product.

3. The more elastic (inelastic) the supplies of productive agents to an industry, other things equal, the more elastic (inelastic) will be the derived supply of the product. Elasticity of supply of a productive agent to an industry depends on the elasticity of the total supply of the agent and on the proportion of the agent employed in the particular industry.

A consideration of these influences as they appear in any particular industry provides the necessary foundation for predicting what the long-run results of an increase or decrease in the demand for the product will be. If the long-run supply of the product is highly elastic big expansions or contractions of output in response to changes of demand will take place and price will change very little. If the long-run supply of the product is very inelastic, responses to changes in demand will be reflected primarily in rising or falling price with only small changes in the quantity produced. In all industries long-run supplies will be more elastic than will short-run supplies; in the long run there is opportunity for more agents to move in or out in response to altered opportunities.

CHAPTER 17

Long-run Demands for Productive Services

A PRELIMINARY analysis of purely competitive demands for productive services was presented in Chapter 15. We are now ready to undertake a more complete analysis. We shall divide our discussion into three main parts: (1) the relation between the demands for a group of products and the derived demand for the services of an agent used in producing them, (2) interrelated demands for different productive services within a firm, and (3) a summary of the major factors influencing the aggregate demand for services of a productive agent.

Derived Demand for a Productive Service

The demand for hosiery workers is derived from the demand of consumers for hose. The demand for carpenters is derived from a wide range of consumer demands—for housing, for products made in plants which are built partly by carpenters, for goods sold in retail stores also built by carpenters. It is because consumers want and will pay for these things that entrepreneurs hire the hosiery workers and the carpenters, that they buy the silk and machinery and steel and coal, the wood and nails and windowpanes. This ultimate dependence of demand for productive agents on consumer demands was implied in studying the short-run behavior of firms in Chapter 15. These firms buy because through using what they buy they can in turn produce goods and obtain revenues from their customers. The greater these consumer demands the greater, other things equal, are these opportunities.

The relation between the demands for a group of products and the demand for the services of an agent contributing to their production is most easily seen by tracing through the processes of adjustment to a change in the supply of the agent. We shall focus our attention on long-run adjustments, paralleling the discussion in Chapter 16.

A long-run derived demand; illustration of adjustment to an increase in the supply of a productive agent

An analysis of adjustments to an increase in the supply of a productive agent could include several steps. We shall look at the immediate results and then at how these lead to the ultimate long-run adjustment. For illustration, we shall suppose that the supply of mill workers (all of

240 LONG-RUN DEMANDS FOR PRODUCTIVE SERVICES

identical qualifications) available to firms in the many branches of the textile industry has increased. In order to simplify analysis we shall suppose these workers to be the only agent that is variable in the short run in any of the firms involved. (One worker is then identical with one unit of input in the short run.)

What will be the immediate, short-run results? Following the analysis of Chapter 15 we should say that more workers would be employed and their wages would fall. Firms hire these mill workers until marginal revenue and marginal cost are equal; under purely competitive conditions marginal revenue is identical with the sale value of the marginal product and marginal cost is identical with average expenditure (wage); hence each firm will employ workers up to the point at which the wage of a

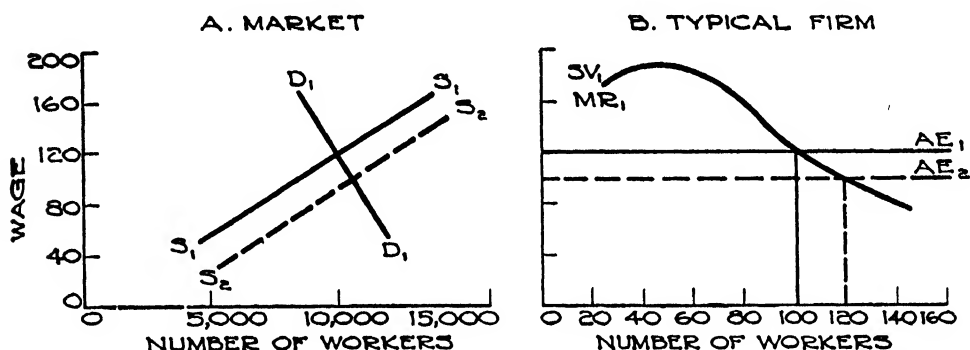


Fig. 17—1. Short-run adjustment to a change in supply of a productive agent.

worker equals the sale value of the marginal product resulting. Workers will undercut each other in order to get jobs, and at the new lower wage each firm will find it advantageous to hire more workers than before.

These short-run adjustments are illustrated in Figures 17—1A and 17—1B. Figure 17—1A gives the aggregate short-run demand and supply schedules for mill workers. The curve S_1S_1 is the initial supply curve; the curve S_2S_2 represents the new increased supply, and the curve D_1D_1 is the initial aggregate short-run demand for these workers. The increase in supply leads to a short-run increase in employment from 10,000 to 10,500 workers, and a drop in wages from \$120 to \$100 a month. The position of a typical firm employing such workers (in any of a number of industries) is illustrated in Figure 17—1B. At the lower wage this firm will increase its employment from 100 to 120 workers. The aggregate increase in mill workers employed will of course be the sum of all these increases in individual firms (jumping from 10,000 to 10,500 workers).

But in each of the industries in which firms hiring mill workers are engaged there will now be profit opportunities (assuming costs were exactly covered before); these profit opportunities appear because of the drop in costs. As a result new firms appear and enter the market

for mill labor. As these new firms become established, there will be a new and greater short-run demand for mill workers; in terms of graphic analysis, the demand for mill workers will have shifted to the right.

So long as firms can enter the industries in which mill workers are employed and more than cover costs there will be an inducement for them to do so. But as more firms appear and demand for the mill labor increases, two forces are brought into operation that tend to reduce and eventually to eliminate the profits earned by these firms. The first of these is an increase of wages resulting from the increase in demand (pulling wages back toward but not to the initial level of \$120). The second is that the prices of the various products produced with this labor will fall as outputs are expanded. When profits in each of these industries

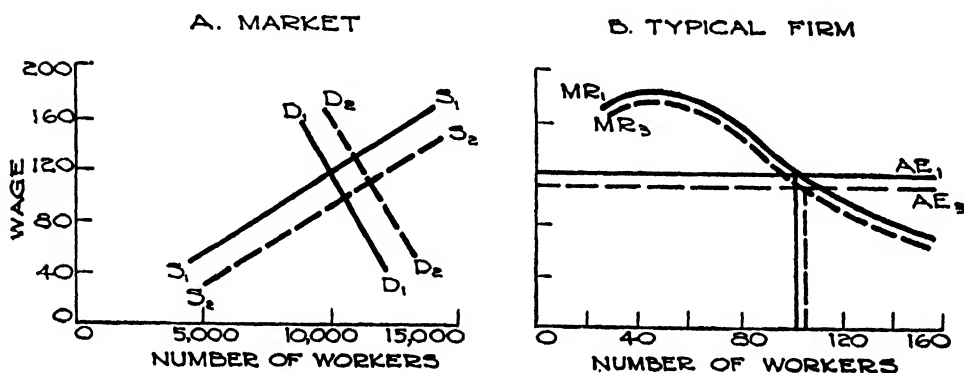


Fig. 17—2. Adjustment after entry of new firms.

have been eliminated once more a new long-run equilibrium position will be attained.

Thus in the long-run adjustment there will be a greater increase in the number of workers hired than in the immediate adjustment when the number of firms remained unchanged. And in the long run the drop of wages will be less than at first. These adjustments are illustrated in Figures 17—2A and 17—2B. Figure 17—2A shows the old and new short-run aggregate demand and supply schedules for mill workers. The demand D_2 represents what firms will take at various wages after there has been time for new firms to enter the field. The ultimate adjustment will be at a level of employment of 11,500 workers for a wage of \$110. Figure 17—2B shows the new adjustment of a typical firm. The curve representing the sale values of the marginal products will have fallen because of the drop in the price of whatever goods are produced with this labor; the new wage (hence marginal cost) will be less than initially, though higher than in the first stage of the adjustment. This firm will ultimately employ 104 workers (which in this particular illustration happens to be just a little more than the number at which it started). The large aggregate increase in employment will be due to

242 LONG-RUN DEMANDS FOR PRODUCTIVE SERVICES

the hiring by new firms that have appeared in response to the lower labor costs.

It is now possible to construct a long-run demand curve for this mill labor. Each long-run adjustment is a short-run adjustment as well; it is a short-run adjustment under conditions in which profits are eliminated. This occurred by assumption when demand was D_1D_1 and supply was S_1S_1 ; it occurred again when demand was D_2D_2 and supply was S_2S_2 . By drawing a line through the points of intersection of these sets of short-run curves we have a long-run demand schedule which is analogous to the long-run supply of a product as presented in Chapter 16. This long-run curve slopes down to the right because of the falling prices of the products made by the mill labor as the outputs of these products are increased.

Relation between elasticities of demands for products and for the services of a productive agent

The products produced by these mill workers might be multiple. In order to simplify our analysis we shall assume that there are only three, all of them being produced in purely competitive industries. Let us say that they are cotton sheets of a certain size and quality, measuring tapes (again all alike), and white cotton gloves. The character of the long-run demand curve for the mill workers will then depend on the character of consumer demands for these three products. If large percentage expansions of output in these industries are accompanied by only very slight percentage drops in prices (that is, if the demands for these products are all highly elastic) the increase in employment of mill workers in the long run will be very great and the drop in their wages will be small. This must be true because with the entry of new firms into these industries and the expansion of outputs the small drops in price bring only slight decreases in schedules of the sale values of marginal products in each of the firms. Profits are then squeezed out primarily by the bidding up again of wages of mill workers with only small drops in the prices of goods sold. In other words; *if the demands for the goods produced by a particular productive agent are highly elastic, the derived demand for that agent will tend to be highly elastic also.*

But it may be that consumers are not so responsive to lower prices in the buying of sheets and measuring-tapes and white cotton gloves; perhaps any appreciable increase in aggregate outputs of these products can be sold only at considerably reduced prices. If the expansion of these industries brings sharp reductions in prices it will bring sharp reductions in the schedules of sale values of marginal products as well. Under such circumstances the increase in employment of mill workers would be small and the drop in their wages would be great. The entry of new firms eliminates profits primarily through the resultant declines in the prices of the goods sold. *If the demands for the goods produced by a particular productive agent are highly inelastic, the derived demand for that agent will tend to be highly inelastic also.*

Interrelated Demands of a Firm for Services of Different Productive Agents

It is now time that we drop the convenient simplification introduced in most of our previous discussions of demands of firms for productive services, that is the assumption that the firm is considering variations of one agent only. Any observation of what happens in a real setting shows that many different kinds of productive agents are generally combined in a firm. The problems of the entrepreneur include many questions as to the proportions in which these agents should be combined. The demands for services of different productive agents that are or may be employed in the same firm are therefore closely interrelated demands.

Characteristics of the most profitable combination of a number of productive agents in a firm

Instead of simply asking, "Shall I employ more mill workers or not?" an entrepreneur in fact asks himself a whole set of questions. Should he put in some machines that would replace some of his unskilled workers but increase the number of mechanics he would hire? With the number of machines of various kinds which he now has would it pay to add another worker? Buy less raw material? Should he drop a worker? Buy less raw material? With the present staff of labor would it be advantageous to get another machine? Should he expand all of these things, his labor force, his machine equipment, his raw material purchases? Some of these decisions involve weighing of one production method against another, and the answer will make a big difference in the kind of productive resources used. But once he has made this basic decision, his problem becomes one of more refined detail. Should he add a little more of this and take off a little of that? All of these are then marginal decisions. *They involve comparisons between the addition to cost (marginal cost) and the addition to revenue (marginal revenue) involved if, starting from any given point, he adds one unit of any particular productive agent. They involve comparisons between the reduction of cost and the reduction of revenue involved if, starting from any given point, he drops off one unit of a particular agent.* Through a series of experiments, partly actually tested out and partly in his imagination, the entrepreneur tries to approximate the most profitable possible adjustment (or the least unprofitable, if that is the best the situation affords).

Let us suppose that a purely competitive firm producing nurses' uniforms buys the following items: labor of stitchers at a wage of \$100 a month, labor of cutters at a wage of \$150 a month, cotton yard goods at \$30 a bolt, sewing machines at \$60 each, and machine tenders at a wage of \$200 a month. We are avoiding a lot of troubles by supposing this to be a short-run situation so that the question of buildings and work space and of many others things can be ignored; and we have of

course omitted dozens of small items that would be purchased in such an enterprise. How many stitchers, and cutters, and machine tenders, how many machines, how many bolts of cotton cloth, would this firm use? Entrepreneurs spend many weeks working out the answers to just such questions, so we shall not attempt a detailed solution here. But we can shortcut the problem of the entrepreneur by describing the nature of the adjustment toward which he will be aiming. Before he arrives at that point he will have considered a great many possibilities. Here would be his final position if he made intelligent choices:

1. The number of stitchers employed will be such that to add another stitcher, given the quantities of other agents he is using, would add less than \$100 to revenue (marginal revenue from the work of another stitcher would be less than \$100); if it would add more than \$100 to revenue he would be able to improve his position by hiring another stitcher, since the addition to cost (marginal cost) would be only \$100. Moreover, to drop a stitcher would reduce revenue by \$100 or more; otherwise (if revenue would be reduced by less than \$100) he could improve his position by curtailing his employment of stitchers, since costs would be reduced more than revenue.

2. The number of cutters employed will be such that to add another cutter, given the quantities of other agents he is using, would add less than \$150 to revenue, and to drop a cutter would reduce revenue by \$150 or more.

3. The number of bolts of cotton cloth used will be such that to add another bolt, given the quantities of other agents he is using, would add less than \$30 to revenue, and to drop a bolt would reduce revenue by \$30 or more.

4. The number of sewing machines used will be such that to add another machine, given the quantities of other agents he is using, would add less than \$60 to revenue, and to drop a machine would reduce revenue by \$60 or more.

5. The number of machine tenders used will be such that to add another machine tender, given the quantities of other agents he is using, would add less than \$200 to revenue, and to drop a machine tender would reduce revenue by \$200 or more.

How may these concrete statements concerning the firm producing nurses' uniforms be generalized to describe the most profitable adjustment of any purely competitive firm? The answer is in fact quite simple: (1) *There will be no further expansion (in the use of any agent) that would add more to revenue (marginal revenue) than to cost (marginal cost); if there were it would have been undertaken.* (2) *There will be no curtailment which would reduce revenue less than it would reduce cost; if there were the curtailment would have been made.* These two statements may be put even more briefly if we assume adjustments can be made in fractions: Inputs of all agents will tend to be adjusted in such a way that for each agent marginal cost and marginal revenue are equal.

If this firm is purely competitive in both selling and buying (our assumption) it will then adjust on all fronts so that the sale value of the marginal product of a unit of a productive agent and the average expenditure on a unit of that agent are equal.

Complementary and substitutive relations between productive agents

There is almost always more than one way of producing a particular product. By some methods relatively unskilled labor may be used extensively and machines very little. By other methods there may be primarily machine production with some skilled labor. Some kinds of machinery may be substituted for other kinds, and some chemical processes for others. The use of any particular machine or chemical process may involve the complementary use of particular cooperating agents, whereas some other process requires the use of other cooperating agents. We see these variations all around us. Houses may be largely prefabricated, machines and unskilled labor being used to a large extent in the factory manufacture of house parts; or similar houses may be constructed largely by hand on the site, using primarily skilled labor and a minimum of machinery. Farms may be mechanized, or they may continue to use horse or mule power. Which will be the "best" method in any particular industry or firm will depend on the prices of the services rendered by the different agents, on their relative technical efficiencies, and, in the short run, on what agents a firm already possesses.

The different relations that may appear as between different productive services in a particular firm may be best understood by the use of an illustration. A concrete example might be the use of farm acreages and labor in agriculture. The farmer may choose between operating a few acres "intensively," or more acres "extensively." If the rent of land is high and wages are low, he may cultivate a small acreage with great care, employing a relatively large number of workers. If land is "cheap" and labor is "high," he may employ fewer workers to cultivate more land with less care. In the latter case he substituted the services of land for the services of labor. Suppose, however, that the drop in land rents makes it worth his while to expand his farm and to hire more labor to cultivate the additional acreage. In this case the demand for services of labor is complementary with that for services of land.

In discussing productive agents, as in the case of consumers' goods, we may distinguish between substitutive and complementary relations. *Two productive services are complementary to each other when an increase in the input of one is associated with an increased demand for the other. Two productive services are substitutive when an increase in the input of one is associated with a decrease in the demand for the other.*

These definitions apply equally to actions within a firm and to the aggregate effects of the actions of the entire group of firms involved. Any two productive services may be "complementary" to each other in some situations and some price relations, while these same agents may be "substitutive" in other situations and price ranges.

Adjustment of a firm to a change in price of one among a number of productive agents

Any change in the price of a productive agent will change the most profitable possible combination of agents in a firm. In order to cut our problem down to manageable dimensions, we shall assume that there are just three kinds of productive agents employed in a particular enterprise, a certain kind of machine, labor, and raw materials. We shall suppose that the labor and the machines can be to some extent substituted for each other: they are in a substitutive relation. And we shall suppose that the raw material stands in a complementary relation with the labor. Now what would happen to the adjustment in this firm if the price of the labor were to change?

Let us suppose that owing to an increase in the supply of labor, wages fall. This happens through the actions of the many buyers and sellers of the labor, but how does it look to any single firm? Since wages of labor are now lower than before, the firm will clearly increase the number of workers employed. With the existing quantity of other agents used in the firm, it would pay to hire more workers, since at the now lower wage they would add more to revenue than they would add to cost. But the adjustment does not stop there. The lower wage will affect the most profitable inputs of both machines and raw materials. Since raw materials are complementary with this labor, the increase in the number of workers will increase the value to the entrepreneur of additional units of raw material and he will therefore buy more of it. He will reduce the number of machines he is using, however, substituting the labor for the machines; with the increased amount of labor used in his firm, the value of a machine has declined, and it will therefore improve his position to curtail somewhat his input of these machines. *When he arrives at his new most profitable adjustment it will again be true that each agent is employed to a point such that, given the existing quantity of other agents in the firm, the sale value of its marginal product equals the amount expended on it per unit (of input).* He will have increased his inputs of labor and of the complementary agent, raw material; he will have diminished his use of the substitute agent, machines.

Suppose, on the other hand, that the wages of labor had risen (owing perhaps to an increased demand for this labor in other industries). Since wages of labor are now higher than before, the firm will probably cut down on its employment of labor; by dropping off workers it will now reduce its costs more than its revenues. There will also be an effect on the employment of raw materials and of machines. Since labor is now so expensive, it will pay to substitute machines for at least some of the labor; each machine is now more valuable to this entrepreneur, since it can be used as a substitute for the labor. With the reduced employment of labor, however, there will tend to be some curtailment of purchases of raw materials; the reduction in labor force has made marginal units of raw material less valuable to the entrepreneur than

before. *Finally, once more he will arrive at an adjustment in which each agent is employed to a point such that, given the existing quantity of other agents in the firm, the sale value of its marginal product equals the amount expended on it per unit. He will have decreased his inputs of labor and of the complementary agent, raw material; he will have increased his use of the substitute agent, machines.*

There is implied in this analysis another generalization that must be made explicit before we shall have a complete statement concerning the characteristics of the most profitable adjustment of a firm. When any two agents can be substituted for each other it might be that another unit of either of them would add more to revenue than to cost and yet the advantage gained by adding one might be greater than the advantage gained by adding the other. By adding either, an entrepreneur might meet the requirements of our previous statements. But an entrepreneur is not in his most profitable adjustment if he could gain by substituting some of one agent for some of another. This is true even if all adjustments between marginal revenues and marginal costs (under pure competition between the sale values of marginal products and the average expenditures on a productive agent) are consistent with the principles already stated. The additional rule is a very simple and obvious one: *In his most profitable adjustment an entrepreneur must have combined agents in such a way that there is no regrouping that would improve his position.*

Summary of Factors Influencing Prices and Aggregate Demands for the Services of a Productive Agent

The threads of discussion in the preceding sections may now be drawn together in a brief conclusion concerning the aggregate demands for the services of productive agents.

1. An increase in consumer demand for a product will increase the "derived" demand for the services of agents used in its production, and vice versa. If the resources (agents) involved are specialized to this particular industry (for example, skilled shoe cutters in the shoe industry), the proportionate effect will be much greater than when the demand in this industry constitutes only a small fraction of the aggregate demand for the resource (for example, the demand of the shoe industry for common labor).

2. The more elastic the demands for the goods toward the production of which a particular agent contributes, the more elastic, other things equal, will be the demand for the services of the agent.

3. When two productive agents are in a substitutive relation to each other, anything which causes the price of one of them to rise will tend to increase the demand for the other and hence to raise its price. Anything which causes the price of one of them to fall will tend to decrease the demand for the other, and hence to lower its price. The more nearly perfect the substitutability, the more elastic the demand for an agent,

248 LONG-RUN DEMANDS FOR PRODUCTIVE SERVICES

since an increase in its price will quickly lead to a shift to inputs of the other agent.

4. When two productive agents are in a complementary relation to each other, anything that causes the price of one of them to rise will tend to decrease the demand for the other and hence to lower its price. Anything that causes the price of one of them to fall will tend to increase the demand for the other and hence to raise its price. This leads demand for each agent to be less elastic than would otherwise be the case, since market adjustments through the lowering of the price of one agent partly compensate for a rise in the price of the other (and vice versa).

Appendix to Chapter 17

Advanced Analysis of Interdependent Demands for Productive Services

Input adjustments assuming two agents variable

Input adjustments of the firm may be clarified by the use of a hypothetical model in which we assume two kinds of productive services to be variable. This may be a long-run situation in which only two kinds of productive services are used; it may on the other hand be a short-run situation in which two variable agents are employed in various combinations with given fixed agents. The basis of the adjustment here, as in the simpler cases analyzed in Chapters 11, 12, 13, and 15, is a comparison of marginal costs with marginal revenues.

Data on total outputs with different combinations of inputs of two productive agents are presented in Table 17—1. Each entry in the table

TABLE 17—1

HYPOTHETICAL TOTAL OUTPUTS; COMBINING SERVICES A AND B

B \ A										
	16	17	18	19	20	21	22	23	24	25
26	3,805	3,904	3,997	4,084	4,165	4,240	4,309	4,372	4,429	4,480
27	3,891	3,995	4,093	4,185	4,271	4,351	4,425	4,493	4,555	4,611
28	3,973	4,081	4,184	4,281	4,372	4,457	4,536	4,609	4,676	4,737
29	4,051	4,163	4,270	4,372	4,468	4,558	4,642	4,720	4,792	4,858
30	4,125	4,241	4,352	4,458	4,559	4,654	4,743	4,826	4,903	4,974
31	4,195	4,315	4,430	4,540	4,645	4,745	4,839	4,927	5,009	5,085
32	4,261	4,385	4,504	4,618	4,727	4,831	4,928	5,022	5,110	5,191
33	4,323	4,451	4,574	4,692	4,805	4,913	5,014	5,110	5,203	5,290
34	4,381	4,513	4,640	4,762	4,879	4,991	5,098	5,194	5,288	5,380
35	4,435	4,571	4,702	4,828	4,949	5,065	5,176	5,276	5,368	5,455

indicates the total output resulting from the associated combination of services A and B. A comparison of two adjacent totals will therefore tell us the difference in total output resulting from a change of one unit in the input of one of these services. For example, when 19 units of A and 29 units of B are combined the resulting output is 4,372. Increasing

input of A by one unit (to 20) results in an output of 4,468, which is 96 more than before. The marginal product of the twentieth unit of A *when 29 units of B are employed* is then 96. The marginal product of the twentieth unit of A when 30 units of B is employed is, however, slightly larger; it is the difference between 4,559 and 4,458, or 101 units. These data are in fact merely a sector of what would be a more complete table of input-output possibilities; they are a part of that range within which entrepreneurial decisions will be made. It is characterized by the fact that, if any given input of A is assumed, say 22 units, each successive unit of B adds a smaller amount to the total product; and assuming any given input of B, each successive unit of A adds a smaller amount to the total product. Marginal products are diminishing.

Now, in order to study the entrepreneur's input policy, let us assume that a unit of service A costs \$95, a unit of service B, \$85. In order to make our analysis as simple as possible, we shall assume further that this entrepreneur is selling in a purely competitive market in which he gets \$1 per unit of output. His total revenue will then be \$1 times his total output, and Table 17—1 may be regarded as a table of total revenues.

This entrepreneur may proceed experimentally to find the best possible combination of services at these costs per unit and under the total revenue conditions given. Suppose he begins experimentally with 24A and 32B. This is clearly not his best possible position. By dropping one unit of A he will decrease his costs by \$95 marginal cost, while his marginal revenue is only \$88 (\$5,110—\$5,023). He will in fact be better off to drop the twenty-third unit of A also, but the twenty-second he will retain at this level since it adds \$97 to revenue (\$4,928—\$4,831) and only \$95 to costs. In this situation B is still underemployed, however. Another unit of B will add \$86 to revenue (\$5,014—\$4,928) and only \$85 to cost. Having added a unit of B, it will now pay to hire another unit of A, adding \$96 to revenue (\$5,110—\$5,014) and only \$95 to cost; it will not now pay to add the thirty-fourth unit of B, since to do so would add \$85 to cost and only \$84 to revenue. This entrepreneur will maximize his profits by the employment of 23 units of A and 33 units of B. If we had assumed A to be priced at \$100 including interest, and B, \$80, in the maximum profit adjustment 22 units of A and 34 of B would be employed. On assuming the price of B to be unchanged when A went up in price to \$100 the employment of both A and B would have been reduced, A to 21, B to 32 units. By assuming the price of B to remain unchanged we may experiment with various prices of A to learn how much of each agent would be taken by this entrepreneur. The best possible adjustment in each case will be such that no addition of an agent will yield a marginal revenue in excess of marginal cost and no curtailment will reduce costs more than it reduces revenues.

In the hypothetical enterprise that we have just examined, the services A and B were combined in various proportions depending on their relative prices. At prices respectively of \$100 and \$85 this entrepreneur would

250 LONG-RUN DEMANDS FOR PRODUCTIVE SERVICES

use 21 units of A and 32 of B. If the price of A were lowered to \$95, B remaining at \$85, the best adjustment would be 23 of A and 33 of B. This would be described in technical language as a *complementary* relation between A and B because a lowering of the price of A resulted in both an increased use by this entrepreneur of A and an increased demand for B. On the other hand, a shift of prices from A at \$95 and B at \$85 to A at \$100 and B at \$80 would result in a *substitution* of B for A, the entrepreneur employing one more unit of B than before, one *less* unit of A. In our example, therefore, the services A and B may be used together in a complementary relationship and they may also be substituted for each other within certain ranges.

We may now ask whether in any given situation with regard to the price of its product and of the agents it employs, this hypothetical firm is making a profit or a loss and whether it will operate in the short run. Let us assume the price of the product to be \$1.00, and the prices of A and B to be \$95 and \$85 respectively.

Suppose, further, that this is a short-run situation and that A and B are the only variable agents employed in this firm. At the best possible adjustment total output will be 5,110 and since it sells at a dollar a unit the total revenue will be \$5,110. Variable cost will be the sum of the expenditure on 23 units of A and 33 units of B, or \$4,990. Revenues exceed variable costs by the differences between \$5,110 and \$4,990 or by \$120. It will pay this firm to operate in the short run. The profit or loss situation depends on what we assume with regard to fixed costs. If fixed costs are greater than \$120 this firm will be operating at a loss and in the long run it will withdraw from the industry unless it can reorganize on a more efficient basis.

On the other hand, if we suppose A and B to be the only agents employed in a new enterprise, total cost in the best adjustment will be \$4,990 and profits will be \$120. This means that there are profit opportunities, and new firms will enter until the falling price of the product and the rising price of productive services eliminate the profit margin.

Interdependence in the pricing of productive services

In constructing the demand of a single firm for a given productive service, A, we assume the prices of products and the prices and the technological efficiencies of other productive services to "remain constant." At various prices for this productive service the firm will take various quantities, and will combine them with different amounts of other productive services that are its complements or substitutes. A first approximation to aggregate demand for A would then be the sum of each of the individual demands of the firms using this service (in whatever industry the firms may operate).

But a serious difficulty immediately arises. This aggregate demand is constructed on the assumption that "other things remain unchanged"; but the shift from one price of A to another will itself cause a change in some of these "other things." The action of the aggregate of firms will

have effects that would not follow from the action of one firm alone. At lower prices for productive service A, the individual demands for productive service B may increase (if it is complementary with A) or decrease (if it is substitutive with A); and vice versa if the price of A were to rise. Therefore the adjustment of each firm to the changed price of A changes the price of B, which in turn changes the demand of the individual firm for A. Let us see how this might work out in some hypothetical market adjustments.

First we may suppose A and B to be complementary services, and that the supply of A for some reason increases. Sellers offering A will undercut each other in order to make sales. As they do this, firms will find it advantageous to take more of A, and they will bid against each other to get more of B to use with A. In bidding against each other for B, the aggregate of firms will bid the price of B up, which discourages somewhat further increases in purchases of A. Thus through a gradual cutting of the price of A and bidding up of the price of B we arrive at a new market adjustment in which a greater quantity of both A and B is taken than before. A higher price is paid for B and a somewhat lower price for A, owing to the increase in its supply.

Now let us suppose that A and C are substitutive services, and that the supply of A for some reason increases. Sellers offering A will again undercut each other in order to make sales. As they do so firms will employ more of A, substituting it in part for C. The more they replace C, the more will be the sellers of C who will be unable to find outlets for their services, and who will therefore start cutting price in order to maintain employment. This will check the process of substitution of A for C. In the new market adjustment there will be more units of A employed than before, fewer units of C, and the prices of both A and C will have dropped.

Thus if A, B, and C are all productive agents employed in the same firms, an initial change in the supply of A will have direct repercussions on the price and employment not only of A but of B and C as well. If the supply of A increases, the result will be increased employment of A and B and decreased employment of C, accompanied by lower prices of A and C and a higher price of B. If the supply of A decreases, the result will be decreased employment of A and B and increased employment of C, accompanied by higher prices of A and C and a lower price of B.

CHAPTER 18

Joint Products

IN JUNE 1940, *Fortune Magazine* published an article entitled "Goodrich Mechanical Rubber . . . Tires are Flattening . . . so Goodrich among others expands its elastic frontier with synthetics and 32,000 products." In the same issue there appears an article on Kresge's, and among the pictures illustrating this article is one of counters piled high with miscellaneous items, the scene entitled "The five and tens expand the circumference of a dime." The next issue features an article on Nylon, one of the thousands of products manufactured by Du Pont. These are some of the things that are making news in business today—new products developed and manufactured by large corporations turning out not just one commodity, but literally thousands. Such diversity is not typical of industry in general, but neither is the one-product organization we have assumed thus far. Even farmer Brown produces more than one commodity, and the corner grocer carries hundreds of different commodities on his shelves.

There are many reasons why these firms produce more than one commodity; these reasons were discussed in some detail in Chapter 5. It is convenient here to remind ourselves of three principal reasons, however. (1) It may be that technical conditions of production make it more advantageous to produce two commodities together rather than separately. In some cases in fact there may be little choice in the matter. One cannot very well butcher a steer, for example, and get only sirloin. In other situations the necessity may be less but the advantages nevertheless great. (2) When two products are usually wanted together by consumers, they will usually be offered for sale together, at least on the retail level, sharing the clerking costs, store rent, and so on, that cannot be allocated to either one or the other alone. (3) By diversifying production in certain ways it is sometimes possible to establish, maintain, and exploit monopoly positions that bring profits to the firm. Monopolistic conditions will be considered in later chapters.

Does the fact that in most cases one firm produces several separate products mean that our analysis of single-product firms has little usefulness in the actual economy? Not at all. The processes of adjustment that we have described apply also in situations in which a firm produces two or more products, but the problem is more complicated and the analysis is therefore also more complex. In other chapters the dis-

cussion runs largely in terms of single-product firms because of the greater simplicity of that analysis, but the same considerations are also in essentials the basis of action in a multiple-product situation. It is the purpose of this chapter to indicate how our analysis of adjustments of a firm producing a single product may be elaborated to describe the case of a multiple-product firm, and to show what effects the actions of individual firms thus described will have on the total market situation for the commodities involved.

Characteristics of "Joint Products"

"Joint products" and "joint costs"

Whenever two or more products are produced by the same firm so that some of the costs of that firm are for both together, they are described as "joint products," and the costs that they share are "joint costs." It would be quite impossible to decide exactly how much of the butcher's wages should be regarded as part of the costs of producing bacon, how much as part of the cost of producing lard. It would be equally impossible to decide how much of the rent paid in a retail store should be considered a cost of retailing gloves, how much of hose. In defining joint products and joint costs we have thus indicated the major factor that differentiates them from single-product situations and complicates the analysis of the behavior of the firm; *some of the costs cannot be separately determined.*

Complementary and substitutive production relations between joint products

A situation analogous to that of complementary and substitutive demands for different productive agents appears in the case of interrelated supplies of joint products. *Complementary joint production between two (or more) commodities, say lard and bacon, arises when an increase in the production of one will lead to an increase in the production of the other (or others) as well. Substitutive joint production between two (or more) commodities, say frying pans and sauce pans, arises when an increase in the production of one will lead to a decrease in the production of the other (or others).* Any two products may sometimes be in one relation and sometimes in the other—this is clearly possible with regard to the frying pans and sauce pans—but in any given situation in the firm one or the other relation will appear. The most extreme case of complementarity arises when a change of a certain amount (say one pound) in the output of one product is inevitably associated with a change of a certain amount (say two pounds) in the output of the other (or others). This may be called *perfect complementarity*; ¹ *the proportions of the two commodities produced are in this case absolutely fixed by technical conditions.*

¹ Perfect complementarity is sometimes described as joint production in "fixed proportions," while any other case—whether complementary or substitutive—is described as joint production in "variable proportions."

Examples of perfect complementarity are hard to find, but in a number of instances the situation is very close to this, especially over limited periods of time. Examples are cotton fiber and cotton seed, wool and mutton, lard and bacon from certain types of hogs within a time span too brief to breed for changing proportions, silver and copper refined from ores in which the two metals appear together. Over longer time periods it is possible to change the proportions of these commodities somewhat, though the relation will still tend to be complementary.

On the other hand, if a furniture manufacturer decides to increase his output of tables, it is quite likely he may find it necessary to decrease his output of desks and vice versa. The materials, the equipment, the factory space, the skilled cabinet-makers, all of these are resources that can be used by him alternatively in producing either tables or desks. If the demand for tables were to increase while that for desks remained unchanged, he would decide it was more worth while to divert some of the resources he previously used in producing desks into the production of more tables. Insofar as this relationship prevails, the tables and desks are substitutive joint products.²

A very interesting case of shifting relation between commodities that have at times been complementary and at other times substitutive in supply is afforded by petroleum products. At first gasoline was a by-product of kerosene; as much kerosene as possible was obtained from the crude oil, and beyond this point any increase in the production of kerosene required more crude oil and resulted in more gasoline. Gasoline and kerosene were so completely complementary in joint production that the proportions between them remained constant as the output of kerosene was expanded. Later demand conditions changed. People wanted more gasoline; and the result was that with increased production of gasoline the output of kerosene tended to diminish. Gasoline and kerosene were substitutive joint products. Now almost as much gasoline as possible is refined out of crude oil, and any increase in the price and output of gasoline results in an associated, complementary, increase in the offerings of kerosene.

Pure competition and the position of joint product firms

The conditions of pure competition among sellers of joint products are the same as among sellers of single products. Each of the products offered is also produced in exactly the same form by many other sellers.

²The most extreme case of variable proportions (which may be either complementary or substitutive) occurs when two or more products of one enterprise are technically completely independent (that is, where changes in the output of one have no necessary connection with the quantity of the other or others). The costs that are common to such products are often called "overhead costs." The retailer who makes many goods available to consumers in the same place and at the same time is an example of a seller of many independent goods with common costs, i.e., "overhead," in the rent of his store and the services of his sales clerks. For present purposes we may regard overhead costs as simply one case of joint costs with variable proportions.

As a result, no single firm is in a position to exert alone any control over market price; profits will be maximized, losses minimized, only by adjustments of outputs to the optimum relationship with given market prices of the products. Thus the individual farmer who produces corn and alfalfa in rotation is unable to alter the market prices of either of these commodities by changing his production plans.

The Output and Pricing of Joint Products That Are Perfectly Complementary

When two products are produced always in the same proportions they may be regarded as just one composite product so far as production adjustments are concerned. There is no possibility of separating any part of the costs of producing one from the costs of producing the other. The same statement holds if there are three, or four, or five such products; but we shall confine attention to examples of two products merely in order to simplify the discussion. What then would be the adjustment in a firm producing two such products, and what would be the prices at which such products would tend to sell? These questions can be answered very simply.

Output policy of the firm

Let us suppose that a firm is mining silver and copper, and that these products are "perfectly complementary" because of technical characteristics of this production. It is impossible to increase the output of one without increasing in the same proportion the output of the other. It may be, for example, that 1 pound of silver always accompanies 2 pounds of copper. Since the production of silver and copper is, by assumption, inseparable and in fixed proportions of 1 pound to 2, we may simply define a unit of output as 1 pound of silver plus 2 pounds of copper. Since they cannot be produced separately or in varying proportions, the products, silver and copper, are essentially one product so far as planning production is concerned. In Table 18—1 we have given a hypothetical

TABLE 18—1

HYPOTHETICAL COSTS OF PRODUCING TWO "PERFECTLY COMPLEMENTARY" COMMODITIES

<i>Composite outputs *</i>	<i>Composite total unit cost</i>	<i>Composite marginal unit costs</i>
21	\$52.00	\$?
22	50.45	18.00
23	49.13	20.00
24	48.33	30.00
25	48.00	40.00
26	48.04	49.00
27	48.48	60.00
28	49.29	71.00
29	50.34	80.00
30	51.66	90.00

* One unit of composite output is 1 pound of silver plus 2 pounds of copper.

schedule of total cost per *composite unit of output* (silver plus copper) and the corresponding *composite marginal unit costs*. We are assuming that outputs can be varied one composite unit at a time merely because this simplifies the analysis.

Any given set of prices of the two commodities may be added together to get the *composite unit price*, the price of 1 pound of silver plus 2 pounds of copper. Under conditions of pure competition this composite unit price will also be the *composite marginal unit revenue* to the firm. Suppose the price of silver to be \$30 a pound, that of copper to be \$10 a pound; the composite unit price would then be \$30 plus twice \$10, or a total of \$50. The addition of one composite unit to output will add \$50 to revenue.

Under these circumstances, what would be the maximum profit output of the firm? The answer is exactly as in the case of a single product, except that in the present case we are comparing composite costs with composite revenues. *Maximum profit output will be the largest at which composite marginal unit revenue is still greater than or equal to composite marginal unit cost, at 26 units in our illustration.* If output were expanded beyond this point the addition to cost (composite marginal unit cost) would be greater than the addition of \$50 to revenue (composite marginal unit revenue). To stop short of this point would be to fail to realize the maximum profit (or minimum loss) position, since the addition of another composite unit would add more to revenue than to costs.

Short-run supply of "perfectly complementary" joint products

Since all production decisions in the case of joint products technically fixed in proportions are based on composite costs and revenues, we might draw up composite supply schedules also. *The number of composite units offered for sale by a competitive producer at each of a series of composite unit prices would be based on the relation between composite marginal unit costs and composite price* (just as in the case of single products the individual supply schedule is based on the comparison of marginal unit costs of the single product and its price). The total amount offered by all firms together at any given composite unit price will of course be the sum of the offerings of individual firms.

Here there is an interesting difference between the single-product situation and that of multiple products in fixed proportions. We cannot draw up separate supply schedules for perfectly complementary joint products. The amounts of silver that will be offered for sale depend not on a relation between costs and the price of silver separately, but on a relation between the composite marginal unit costs of silver and copper together and the prices of silver and copper together. The amounts of silver and copper offered at any given composite unit price will be the same whether the price of silver is high and copper low or vice versa, so long as they add up to the same amount for the two together. Thus if silver were \$20 a pound and copper \$15, so that one composite unit of silver plus copper were (\$20 plus twice \$15) still \$50, the adjustment would be exactly the same as when the price of silver is \$30 and that of copper \$10.

Demand schedules for joint products

The demand schedule of a particular commodity is usually constructed on the assumption that the prices of other commodities remain unchanged. Strictly then, we should construct different demand schedules for silver at each assumed price of copper and vice versa. Another alternative would be to assume that the demands for silver and copper were completely independent of each other so that whatever the price of silver the demand schedule for copper would be unchanged; whatever the price of copper the demand schedule for silver would be unchanged.

On this assumption that the demands for silver and for copper are independent of each other, we may construct their separate demand schedules. Such hypothetical schedules are given in Table 18—2 and

TABLE 18—2
HYPOTHETICAL DEMAND SCHEDULE FOR SILVER

<i>Quantity of silver (pounds)</i>	<i>Price of silver</i>
50,000	\$40
100,000	35
150,000	30
200,000	25
250,000	20
300,000	15

Table 18—3. By adding together the associated prices of silver and

TABLE 18—3
HYPOTHETICAL DEMAND SCHEDULE FOR COPPER

<i>Quantity of copper (pounds)</i>	<i>Price of copper</i>
100,000	\$12
200,000	11
300,000	10
400,000	9
500,000	8
600,000	7

copper we arrive at the composite unit prices at which each quantity of composite units of silver and copper would be taken by consumers. This composite demand schedule is given in Table 18—4. For example, when

TABLE 18—4
HYPOTHETICAL COMPOSITE DEMAND SCHEDULE; SILVER PLUS COPPER

<i>Quantity of composite units *</i>	<i>Composite unit price</i>
50,000	\$64
100,000	57
150,000	50
200,000	43
250,000	36
300,000	29

* One composite unit equals one pound silver plus two pounds copper.

50,000 pounds of silver are produced there will be an associated output of 100,000 pounds of copper; 50,000 pounds of silver will be purchased at \$40 a pound, and 100,000 pounds of copper will be purchased at \$12 a pound. Since a composite unit of output has been defined as one pound of silver plus two pounds of copper, 50,000 pounds of silver plus 100,000 pounds of copper make up 50,000 composite units of the two products together. The composite unit price at this output level is \$40 plus \$12 plus \$12, or the \$64 entered in Table 18—4; that is, one pound of silver plus two pounds of copper will sell for a total of \$64 when the output of these metals is 50,000 composite units. Each of the other items in Table 18—4 is similarly derived from the separate demands for silver and copper.³

The short-run pricing of "perfectly complementary" joint products

The pricing process in all competitive markets is a process of competition between sellers to sell their goods, between consumers to buy these goods.⁴

Suppose that the total offerings of all firms together were as indicated in Table 18—5. This is a hypothetical composite short-run supply schedule for the commodities silver and copper, the sum of individual supply schedules based on the relations between composite marginal unit costs and composite unit prices. Under the demand conditions assumed by Tables 18—2 and 18—3, and by the composite table derived from them, Table 18—4, the composite market price would tend to the level \$50, with an output of the industry of 150,000 composite units (that is 150,000

³ Unfortunately, at least for the student of economics, the assumption that the demands for two joint products are mutually independent is not always justified. While it may be a reasonable assumption in the case of wool and mutton, for example, it is clearly not a reasonable assumption in the case of different cuts of meat from the same animal. Interdependence of demand does not create any serious problem in the analysis of single product industries, since shifts in production responding to changes in demand for one product do not change automatically the "other prices" assumed unchanged in the construction of demand schedules. Such interdependence cannot be so lightly dismissed in the case of joint products, where shifts in outputs of one product are directly associated with shifts in outputs and hence prices of other products. Given any particular underlying set of consumer preferences, there are many different price and quantity relations as expressed in the various demand possibilities for two products closely related in consumers' choices. The construction of various demand possibilities under these circumstances involves analysis too complex for us to undertake here. It is important, however, that we recognize that the existence of such interrelation may have important implications for social policy. More than one error has been precipitated by attempts to apply the simplified analysis of a single product case to a case of joint products where demands are interdependent. Nowhere have such errors been more striking than in analysis of the effects of excise taxes on certain commodities that are characterized by these complex interrelationships on the side of both demand and supply.

⁴ In some cases this process appears in very complex form; in some cases it is relatively simple. We shall not undertake here to follow through pricing of products when the demands for these products are interrelated, since this analysis would be far too complex. It is useful, however, to consider the process as it would function under the assumption that these demands are mutually independent.

TABLE 18—5

HYPOTHETICAL COMPOSITE SHORT-RUN SUPPLY

<i>Quantity of composite units *</i>	<i>Composite unit price</i>
50,000	\$30
100,000	40
150,000	50
200,000	60
250,000	70
300,000	80

* One composite unit equals one pound silver plus two pounds copper.

pounds of silver plus 300,000 pounds of copper). At any composite price higher than this more would be offered than was wanted; at any price lower more would be wanted than was offered. Just as in the case of a single product, *the competition of buyers and sellers would lead to an adjustment at a price for each product such that the amount wanted and the amount offered would be the same.*

What determines the separate prices of silver and of copper? Their separate prices will depend on the shapes and positions of the separate demand schedules, as presented in Tables 18—2 and 18—3. Since the demand for silver is great relative to that for copper, the price of silver is higher than is that for copper. Having found the short-run equilibrium adjustment of the industry for the two products together, we can find the separate prices of each by reading from their separate demand schedules the prices at which the equilibrium quantities of each could be sold: 150,000 pounds of silver will be purchased at \$30 a pound (Table 18—2); 300,000 pounds of copper will be purchased at \$10 a pound (Table 18—3).

Long-run adjustments

It is evident that the industry of which the firm in our preceding discussion is a member would not be in a position of long-run equilibrium. At the optimum output of 26 composite units, total revenue of this firm would be 26 times \$50, or \$1,300; total cost would be 26 times 48.04, or \$1,093; and profit would be \$1,300 — \$1,093, or \$207. Anticipating profit opportunities, new firms would enter this industry,⁵ and they would continue to enter until there were no further profit opportunities, until the composite unit price was just equal to the lowest composite total unit cost.

What of the separate prices of silver and copper in this long-run adjustment? Their relative prices will still depend on the relative positions and shapes of the separate demand schedules. As new firms appear in the industry and outputs are increased, the prices of both products will fall, but one may fall much more than the other, depending on the degree of demand elasticity for the product. If the demand for silver were very elastic and the demand for copper were very inelastic, the new adjustment

⁵ Assuming freedom of entry.

would involve a much sharper drop in the price of copper than in that of silver.

How does this analysis compare with that of long-run adjustments in single-product industries? It is essentially the same. New firms will enter an industry in which profits are anticipated; firms will withdraw when they anticipate that losses would result from reinvestment. *In the long-run equilibrium position lowest composite total unit cost will equal the composite unit price. The only difference lies in the impossibility of separating out the costs of the two joint products and comparing these separate costs with separate prices.*

The Output and Pricing of Joint Products the Proportions of Which Can Be Varied

While there may be many instances of almost perfect complementarity in short-run adjustments, over longer periods it is almost without exception possible to vary proportions somewhat. We are now ready to examine these cases, assuming either complementary relations that are only partial (not perfect) or substitutive relations.

Output policy of the firm

The problems presented by production of joint products that can be varied in proportions are almost exactly like those involved in discussing interdependent demands for the services of productive agents in Chapter 17. Let us assume a manufacturer of kitchen furniture to be producing just three standardized items that are exactly like the furniture made in hundreds of other firms; these items might be straight wooden chairs, ladder footstools, and kitchen tables. The chairs sell for \$2.50 each, the footstools for \$3.00 each, and the tables for \$3.50 each, at current market prices. This producer will seek to arrive at the most profitable possible combination of outputs of chairs, footstools, and tables; and he will seek to find this most profitable possible combination by both actual and mental experimentation. When he arrives at the most profitable (or at the least unprofitable) position it will have the following characteristics:

1. The number of chairs produced will be such that to add another chair, given the quantities of tables and footstools he is producing, would add more than \$2.50 to costs (the marginal unit cost of producing another chair would be more than the resulting marginal unit revenue of \$2.50); if it would add less than \$2.50 to cost he would be able to improve his position by adding another chair, since the addition to revenue would be \$2.50. Moreover, to produce one less chair would reduce costs by less than \$2.50; otherwise (if costs would be reduced by more than \$2.50) he could improve his position by curtailing his output of chairs, since costs would be reduced more than revenue.

2. The number of footstools produced will be such that to add another footstool, given the quantity of chairs and tables he is producing, would

add more than \$3.00 to costs, and to produce one less footstool would reduce costs by less than \$3.00.

3. The number of tables produced will be such that to add another table, given the quantity of chairs and footstools he is producing, would add more than \$3.50 to costs, and to produce one less table would reduce costs by less than \$3.50.

These findings may be summarized under a general statement to describe the most profitable adjustment of any competitive producer of joint products the proportions of which can be varied. (1) *There will be no further expansion in the output of any product that would add less to cost than to revenue; if there were it would be undertaken.* (2) *There will be no curtailment of output of any product that would reduce cost more than it would reduce revenue; if there were, the curtailment would be made.* These two statements may be put even more briefly: outputs of all products will tend to be adjusted in such a way that for each product marginal unit cost and marginal unit revenue are equal. If this firm is selling in purely competitive markets (our assumption) it then follows that *adjustments on all fronts will tend to be such that for each product marginal unit cost and selling price will be equal.*

Adjustment of a firm to a change in the price of one among several joint products

Any change in the price of one of a group of joint products will change the most profitable combination of products in a firm. Let us suppose that in our furniture factory the production of chairs and of tables is largely substitutational while the production of chairs and footstools is largely complementary.⁶ Now what would happen to adjustments in this firm if the price of chairs were to change?

First, let us suppose that due to an increase in the demand for chairs the price obtainable rises. Since he can now get more for chairs than before, this manufacturer will be likely to produce more of them; but that is not all. The rise in the price of chairs will affect his policies with regard to footstools and tables as well. Since footstools are complementary with chairs, the increase in the output of chairs will mean that additions to costs in expanding the output of footstools will be less than would previously have been the case, and he will therefore expand his output of footstools. He will reduce the number of tables he is producing, however, diverting the wood and labor and other resources to producing more chairs and footstools. *When he arrives at his new most profitable adjustment it will again be true that each commodity is produced to a point such that, given the existing quantities of other products in the firm, its marginal unit cost and selling price are equal. He will have increased*

⁶ It is important to recognize here that the substitutive and complementary relations discussed are with regard to the production conditions in firms, not with regard to consumer demands for these products; relations in consumer demands might be quite different from those arising from technical factors in production adjustments.

his production of chairs and of the complementary good, footstools; he will have curtailed his production of the substitute good, tables.

Now let us suppose that the price of chairs, instead of rising, had fallen. This will discourage the production of chairs; price is not as great as marginal unit cost at the previous output adjustment. The reduction in output of chairs will lead to a reduction in output of the complementary product, footstools, as well. But this will encourage a substitution of tables for chairs and footstools in the activities of this furniture manufacturer; the output of tables will be increased. *Finally, once more he will arrive at an adjustment in which each commodity is produced to a point such that, given the existing quantities of other products in the firm, its marginal unit cost and selling price are equal. He will have decreased his production of chairs and of the complementary good, footstools; he will have increased his production of the substitute good, tables.*

There is implied in this analysis (as in the parallel analysis of input adjustments in Chapter 17) another generalization that must be stated in order to complete our characterization of the adjustment of outputs of joint products. When any two products can be substituted for each other in the production processes of the firm it may be that another unit of either of them would add less to cost than to revenue, and yet the advantage gained by adding one may be greater than the advantage gained by adding the other. In his most profitable adjustment the entrepreneur must have adjusted his outputs of different goods in such a way that there is no regrouping that would improve his position.

The supplies and prices of joint products

From the previous discussion it is evident that it would be quite impossible to construct a supply schedule for any one among a group of joint products without taking into account the prices at which the other products would sell. The output of tables in our hypothetical firm depended on the prices not only of tables, but of chairs and footstools as well. This is not, however, a counsel of despair. There are some important conclusions which may be drawn from this analysis:

1. When two joint products are in a substitutive relation to each other, anything that causes the price of one of them to rise will tend to decrease the offerings of the other and hence to raise its price. Anything which causes the price of one of them to fall will tend to increase the offerings of the other and hence to lower its price. *The more nearly perfect the substitutability, the more elastic the supply of each product, since a drop in the price of one product will lead producers to shift quickly to the other and vice versa.*

2. When two joint products are in a complementary relation to each other, anything which causes the price of one of them to rise will tend to increase the offerings of the other and hence to lower its price. Anything which causes the price of one of them to fall will tend to decrease the offerings of the other and hence to raise its price. *The more nearly*

perfect the complementarity the less elastic the supply of each product, since market adjustments through the lowering of the price of one product will partly compensate for a rise in the price of the other (and vice versa).

The market adjustments in the pricing of these joint products will again tend to fulfill the essential conditions characteristic of single products. When the market is in equilibrium (either short-run or long-run), *all firms will be operating at output combinations such that the separate marginal unit costs of each of the products are equal to (or just less than) their separate selling prices.* Under these circumstances no seller can gain in the short run by any readjustment available to him. *The competition of sellers and buyers will have resulted in an adjustment such that the amount of each product wanted at its market price equals the amount offered at that price, given the existing prices for the other products.*

Long-run adjustments

All long-run adjustments involve the entry or withdrawal of firms in response to profit (or loss) anticipations. If entrepreneurs see profit opportunities in producing joint products, new firms will be established. If losses are anticipated, existing investments will not be replaced. But although long-run adjustments are like those in single product situations in that new firms will appear when there are profit opportunities, and old firms will withdraw when losses are anticipated, there is one significant difference. It is impossible to separate completely costs of the two or more products produced jointly. Therefore, it is impossible to speak of separate total unit costs, and to compare them with prices. Moreover, we cannot evade this difficulty by speaking of composite total unit costs, if the proportions of the products can be varied, since there is no such thing as a "composite unit."

Significance of Joint Products in the Use of Single Product Analysis

In the preceding pages we have applied marginal analysis, and the analysis of long-run adjustments to profit anticipations, to cases of joint products. There is nothing new in the type of analysis which we have employed, but new complexities appear in the adjustments examined. Some of these complexities are such that firms will inevitably make much rougher approximations to the maximum profit positions described than would be expected where the problems of adjustment were simpler. We are now ready to review these findings:

1. Certain generalizations apply to all cases of pure competition:
 - (a) Outputs of products in each firm will tend to be adjusted to a level such that marginal unit cost and selling price are equal; in the special case of perfectly complementary joint products marginal costs cannot be determined separately and the comparison is therefore between composite marginal unit costs and composite unit price.
 - (b) Market prices (in both short- and long-run adjustments) will tend to levels such that, with the

existing prices for all other products, the price of each exactly equates the amounts offered by sellers and the amounts wanted by the buyers.

(c) Wherever there are profits or losses, firms will tend to move in or out until costs are just covered.

2. The most important differences between single-product analysis and the joint-product situations are these: (a) Total costs cannot be separately determined for joint products, so that comparisons between price and total unit cost are not very meaningful in such cases (though composite total unit cost can be compared with composite unit price in the case of perfectly complementary joint products); the alternative is to turn to a comparison between total revenues and total costs of the firm in its most advantageous adjustment. (b) The supplies of joint products are mutually interdependent and cannot be considered separately except under very specific assumptions concerning the prices of all other products in the joint group; this is in fact merely a refinement of the assumption made in dealing with single products that "all other things remain unchanged." (c) The relations between joint products have important effects on the degrees of responsiveness of sellers to changes in the price of any one of the products; they lead sellers to be more responsive (supply to be more elastic) where the relation is substitutive, less responsive (supply to be more inelastic) where the relation is complementary.⁷

Do these differences invalidate the use of the simplified models of single-product analysis to explain basic processes in a free enterprise economy in which multiple products are the typical situation? The answer is definitely, No. The basic similarities are for most purposes more important than the differences. Single-product analysis will give us the right answer with regard to most questions we might ask about products that are in fact produced jointly with other goods. It is only when we are especially interested in adjustments among closely related products that the further refinements of joint-product analysis are needed. Such refinement is important, for example, if we wish to study the ramifications of a tax on sirloin as it would affect other beef products, the effects of a tariff or subsidy to wool producers on the sales and prices of mutton, and other similar problems.

Appendix to Chapter 18

Advanced Analysis of Output Adjustments of Joint Products the Proportions of Which Can Be Varied

Suppose the products A and B to be variable in proportions, and the total costs to the firm of the various output combinations of these two

⁷ Moreover, when we attempt to carry the analysis of the behavior of individual firms into an analysis of market positions, we run into a whole new set of considerations in the interdependence of products on the side of demand. While interdependence of demand is a very common fact throughout the economy, it introduces no serious difficulties of analysis except where it is associated at the same time with interdependence in production.

products to be as indicated in Table 18—6. How should we read this table? Each column shows the total costs of producing various outputs of B in combination with some given output of A. Thus column 1 of the table gives the total costs of producing 17 units of A plus 28 of B, 17 units of A plus 29 of B, 17 units of A plus 30 of B, and so on. By comparing these successive total costs we may determine the marginal unit cost of, say, the twenty-ninth unit of B when produced in combination with 17 units of A. Column 2 gives us similar information when 18 units of A are being produced. The marginal unit cost of adding a twenty-ninth unit of B will vary, depending on the proportions in which B was previously produced relative to A. When there are 17 units of A produced, the twenty-

TABLE 18—6

HYPOTHETICAL TOTAL COSTS OF PRODUCING TWO COMMODITIES
IN VARIABLE PROPORTIONS

Units of Product B	Units of Product A			
	17	18	19	20
28	\$167.50	\$172.50	\$178.50	\$185.50
29	170.00	174.50	180.00	186.50
30	173.00	177.00	182.00	188.00
31	176.50	179.70	184.20	189.90
32	180.50	183.50	187.00	192.70

ninth unit of B would add $\$170.00 - \167.50 , or $\$2.50$ to costs; when 18 units of A are being produced, the addition of a twenty-ninth unit of B would add $\$174.50 - \172.50 , or $\$2.00$ to costs.

In the same way we may determine marginal unit cost of, say, the eighteenth unit of A when a given quantity of B is produced. Reading across the table horizontally, we find the total cost figures for producing 17, 18, 19, 20 units of A in combination with 28 units of B; or, in the second line, the total costs of producing 17, 18, 19, 20 units of A in combination with 29 units of B, and so on. By comparing successive totals in this direction, we arrive at figures for the marginal unit costs of additional units of A *when produced in combination with some given quantity of B*.

Thus the table of total costs of different output combinations of A and B serves as a basis for estimating the change in costs that would result from changing the combination of outputs by adding or subtracting one unit of either commodity. We are able to arrive at separate marginal unit cost figures, not merely the composite figures obtained when production was in fixed proportions.

How will the producer whose costs are given in Table 18—6 arrive at his optimum output adjustment? To answer this question, we would have to know his expectations with regard to the separate prices of A and of B

We shall suppose the price of A to be \$4.75, that of B \$2.85. Any addition of a unit of A will then add \$4.75 to his revenues; any decrease of one unit of A will diminish his revenues by \$4.75. Similarly, any addition of one unit of B will add \$2.85 to his revenues; any decrease of one unit of B will diminish his revenues by \$2.85. He will compare these separate marginal unit revenues with the various separate marginal unit costs.

Let us experiment with the problem confronting this enterprise. Suppose we start at random at any output combination, say 19 units of A and 30 units of B. Will it pay to make any changes in this output combination, or have we already happened on the best possible position? We shall certainly not increase our output of A by another unit, since to do so would add \$6.00 to costs and only \$4.75 to revenue. In fact, given 30 units of B we are already producing too much of A. If we drop off the nineteenth unit we shall lower our costs by \$5.00 (the marginal cost of the nineteenth unit of A produced in combination with 30 units of B is \$5.00). Dropping off the nineteenth unit will then lower costs more than the loss of \$4.75 in gross revenue. The eighteenth unit will, however, be retained, since it adds but \$4.00 to costs and \$4.75 to income. But it would pay now to increase the output of B to 31 units, since this will add but \$2.70 to costs, \$2.85 to revenue. And with 31 units of B it will pay to increase output of A to 19 units again, adding \$4.50 to cost, \$4.75 to revenue. This is an optimum position since it will not pay to change the output of either product. Stated in more general terms, *a firm will continue to add units of each product so long as to do so will add less to costs than to revenue*; it will not add units that add more to cost than to revenue. This is exactly the kind of adjustment made in single-product firms.

The determination of the optimum output positions of firms producing joint products in variable proportions is thus possible for each combination of prices of the products. If we had assumed prices of \$4.50 for A and \$3.10 for B, the optimum adjustment would be found in the same way, and it would be at a different combination than with prices of \$4.75 and \$2.85. This understanding of the adjustment of individual firms again gives us a basis for constructing individual supplies and hence aggregate supplies.

This analysis shows clearly what the reader should have suspected all along. Any two or more commodities may be in a complementary relation for some ranges of ratios between their prices and in a substitutive relation for others. This fact would have to be taken into account especially where extreme shifts in price are involved, and where variations of proportions are easily accomplished.

CHAPTER 19

Purely Competitive General Equilibrium

IN THE last few chapters we have been gradually "working ourselves up" to the point where we could conceive of and describe such a thing as a general purely competitive equilibrium of an economy as a whole. More and more elements of interdependence in the economic system have become apparent—interdependence between demands for various products and for productive agents used in making them, interdependence between supplies of products and the supplies of resources used in their production (along with demands for these same resources in other uses), interdependence in the buying of different resources, interdependence in the producing of various goods. Changes introduced at any one point vent their force in all directions through the economic system. Before there has been time for a full adjustment to any one change new situations arise and a new wave of adjustments spreads through the economy. To analyze all these many things at once would be a hopeless task; and to analyze them a bit at a time and stop there is to leave the puzzle in pieces. We can, however, pull these parts together in a description of what the characteristics of a purely competitive economy would be if all the forces generated at a given time could be followed through to an ultimate stable position without the intervention of new conditions giving rise to new sets of adjustments. A description of this imaginary purely competitive general equilibrium provides us with a sort of guiding North Star where otherwise the complexity of economic relations would seem to be only darkness and confusion.

The Characteristics of General Equilibrium in a Purely Competitive Economy

There is probably no more abstract concept in economics than that of an economy purely competitive throughout and in complete general equilibrium. Yet this concept has been used extensively and is a convenient tool for many purposes. What are the conditions that would prevail if all adjustments in a purely competitive economy were given time to "work themselves out" to a general equilibrium? This general equilibrium would be attained when there remained no opportunity for any individual or firm to improve his position by any change in his activities or role in the economy. Its characteristics would be as follows:

1. In each firm productive agents would be hired in a quantity and combination such that: (a) No further addition of a unit of any agent to input would result in a marginal sale value greater than average expenditure. (b) No further curtailment of input would involve a reduction of cost greater than the associated reduction of revenue. (c) No alternative method of production would improve the position of the firm.

2. In each firm joint products would be produced in a quantity and combination such that: (a) No further addition to the output of any product would result in a marginal unit cost less than selling price. (b) No further curtailment of output would involve a reduction of cost greater than the associated reduction of revenue. (c) No alternative combination of product outputs would improve the position of the firm.

3. No firm would be making profits or incurring loss. This follows because so long as profits are available in any industry firms will move into these areas with a resulting downward pressure on price and upward pressure on costs until profits are eliminated; and conversely so long as there are losses firms will drop out until rising prices and falling costs restore a situation in which there are neither profits nor losses. Wherever there are either profits or losses there are unstable conditions and incentives to change that result in the elimination of the profits or the losses as the case may be.

4. All consumers who are willing to pay at least the existing price for a particular good are satisfied, and no one would improve his position by buying more of one thing and less of another.¹

5. All sellers of productive services who are willing to offer their services at the existing price or less are satisfied, and no one would improve his position by changing the use to which his services (whether of labor or of the material productive agents he owns) are put. Each seller of a productive service is receiving a return equal to the sale value of the marginal product of that service in its most productive possible employment.

Practical Usefulness of the General Equilibrium Concept

The analysis of long-run adjustments toward a general equilibrium position such as we have outlined has great practical usefulness; it also has important limitations. It is necessary to recognize the possibilities and the limitations in the application of such analysis if we are to make the most effective use of it.

Several important assumptions underly the preceding analysis of general equilibrium:

1. First is the assumption that entrepreneurs act to maximize profits. To the extent that this assumption is unrealized, there will be departures from the adjustments we have described. Lethargy and preferences for

¹ The marginal worth of a dollar spent on one good is equal to the marginal worth of a dollar spent on any other good.

an activity in which an entrepreneur is already engaged are frequently important checks to the withdrawal of firms in declining industries. This is especially important when the nature of the industry definitely identifies it with a way of life, as in agriculture; and it explains some of the lags or failures to adjust that are sometimes regarded as inexplicable.

2. For the most part we have assumed that entrepreneurs were sufficiently correct in their anticipations of costs and prices so that there was a tendency to adjust to those positions that would realize their goals of maximizing profits. The very fact of the appearance of profits and losses is evidence that these anticipations are not always correct, but the discussion of the direction of readjustment in the light of such profits and losses implies that results will be sufficiently successful, typically, to lead at any particular time toward the realization of a long-run equilibrium position.

3. Throughout our analysis we have assumed pure competition with freedom of entry. Such a situation appears only when there are many firms buying or selling the same thing under the same conditions, and it does not always appear even then. If large groups of entrepreneurs find ways of acting in collaboration privately, as in the development of some kinds of producers' cooperatives for the sale of their products, or if they develop political power groups strong enough to bring about government intervention in the operation of the industry, then there will no longer be pure competition despite the large number of firms buying and selling the same things. As large groups find themselves in less and less favorable positions such collaboration and political action becomes increasingly likely, since the common interest and the advantages of group action become more evident to each of the individuals involved. Moreover, wherever firms are differentiated in buying or selling, wherever there are only a few firms, and wherever those who deal with firms are few or differentiated, there will be monopolistic or monopsonistic situations. Thus, though the analytical model of pure competition with freedom of entry is approximated in reality in the production of a wide range of products, it is seldom completely true and over a large portion of our economy it is completely inapplicable. It is therefore important that competitive equilibrium analysis be applied with care and discrimination.

4. In following through processes of adjustment we have ignored the fact that new developments intervene all along the way, changing the position of the ultimate equilibrium before there has been time to complete the adjustment process under any one set of conditions. This might be regarded more as a simplification of analysis than as an assumption: we are studying only one set of adjustments at a time. The prevalence of change is, however, a sufficient reason for expecting that a position of long-run equilibrium would never actually be attained. Consumers' demands do not long remain constant; the amount and composition of available resources is always changing; shifts of technology come frequently in almost every industry. On the basis of conditions existing today adjustments toward a certain long-run equilibrium might be started. but by

tomorrow some change in conditions may require a reversal of these adjustments. By tomorrow consumers' demands may have changed, as a result of changed preferences, or perhaps changed incomes. Or perhaps a new invention appears that drastically lowers costs. Or a new deposit of mineral raw materials may be discovered, changing the relative costs of the various resources used. Or workers may change their preferences in jobs. There is little that remains fixed for long, and never would all relevant factors remain constant for long enough for the processes of long-run adjustment to work themselves out completely toward one long-run position.

To recognize that a general equilibrium position is never in fact realized, and that many of the assumptions on which such analysis is based are only partially applicable to concrete situations, is not to deny the value of such equilibrium analysis. Its purpose is to show the direction of adjustment and the processes of adjustment. There is no intent to give through such analysis a realistic cross-section picture of the economy at any given time—the economy is not like that. But such analysis does have two distinct uses:

1. It provides an understanding of the direction of adjustments and the processes of adjustment that tend to occur in an economy insofar as the assumptions just listed are applicable.
2. It supplies a starting point for evaluating the significance of departure from the assumptions on which it is based. By understanding the model of purely competitive general equilibrium we are better equipped to understand what happens when entrepreneurs act contrary to the profit motive, when they make errors, when there are interferences with pure competition. Through such an approach we shall gain further understanding of the fortunes of various industries and individuals, and of the uses to which the resources of the economy are put.

CHAPTER 20

The Consumer Looks at Pure Competition

THE essential function of the economic system is to direct the use of resources into the production of goods as desired by the people in control. The people in control may be a minority group dictating the ways in which resources should be used for themselves and for the rest of the members of the society as well. The people in control may be the many individual consumers who determine by the sum of their actions the ways in which resources will be used. The people in control may be the elected representatives of a free people who vote to use the collective resources in the building of roads or schools or battleships. Here we are interested in Tom and John and Helen and Mary. We want to know whether the resources of the economy are to be used as these individual consumers would prefer that they be used. We cannot yet answer a question so broad as "Are resources in the American economy used consistently with the preferences of all the individual consumers scattered all over the United States?," though a partial answer has already been undertaken in a preliminary form in Chapters 11 and 12. We are now ready to take the next step. By examining in detail what would happen to resource allocation in an imaginary situation of a purely competitive economy, we take that step. Such an economy would do about as good a job of allocating resources according to consumer preferences as would any imaginable economy;¹ it therefore provides a kind of marker. In later chapters more realistic types of situations may be evaluated by comparing the resource allocation that occurs under those circumstances with allocation under the assumed conditions of pure competition.

What Are Consumer Preferences?

If we are to judge how good or poor a job is done by the economic system in using its resources consistently with consumer preferences, we must have some way of measuring those preferences. Two major problems immediately arise: (1) How are we going to decide whether Tom's wish for another pair of socks is more or less important than John's desire

¹ The Democratic Socialism of Oscar Lange might be cited as a superior conception, though there are many who would regard it as politically "unimaginable."

for a baseball bat, or Mary's desire for a pair of gloves? (2) How are we going to decide whether Tom prefers five pairs of shoes and twenty shirts, or six pairs of shoes and seventeen shirts; whether he prefers to consume ten coca colas a month and to attend one movie or to consume five coca colas and attend two movies; whether he would rather have a slightly higher quality overcoat and a cheaper radio or a better radio and a cheaper overcoat? In other words, how are we going to decide what the preferences of each individual consumer are?

Both of these questions will have to be answered before we can evaluate the performance of the economy in responding to these consumer preferences.

The division of consumer votes

The first of the questions posed above is the most unanswerable, and therefore, paradoxically, the most easily answered. We simply accept as a measure of the relative importance of Mary's want and of Dick's want and of Tom's want the number of dollars each is willing to spend to satisfy his wishes. Perhaps this is not a fair way of judging; most people would probably say it was far from ideal. The rich man can always pay more for a thing than can a poor man. The rich man will get some of the things the poor man would have liked to have, perhaps things he wanted more than did the rich man who got them. The rich man can also spend his money on goods that would not be produced at all were he not there to buy them; it is his expenditures that explain the production of many of the things advertised in the pages of the *New Yorker*. It is by the spending of dollars, in fact, that individuals in the United States normally make their bids for goods. The power of each consumer to direct the use of resources is largely measured by the amount of income at his command. We therefore dispose of the question of deciding whose wants are the most important by simply waiving it aside. We say that "consumers' powers to claim goods are divided according to their incomes; this is just the way things are" (though not the way they necessarily have to be, and we can of course change those incomes). And then we go on to re-define the meaning of getting resources allocated in accordance with consumer preferences. We are in fact going to talk about whether resources get allocated in accordance with consumer preferences *when each consumer has so many dollar votes in the determination of that allocation*.

The expression of individual preferences

How much is a pair of shoes worth to John? Is it worth \$5 or \$8 or \$15? That will depend on how many pairs of shoes and other things he already has as well as on his income and his scale of preferences. And if we find that the fifth pair of shoes is worth \$10 to John, we might then ask what is \$10 worth? And the only way we could answer that question would be to see what he would otherwise spend that \$10 on—perhaps some theatre tickets or books or more clothes, and so on. And

how much \$10 in books is worth to John will again depend in part on how many books he already has. In other words, we measure the worth of something in terms of the things we must forego to get that thing; and when we consider how much we would give up to get a fifth pair of shoes we consider not what we would sacrifice to get 1 pair of shoes when we had none, but rather what we would sacrifice when we already had 4.

John will probably be willing to give up more roast beef to have a third pair of shoes than he would sacrifice for a fourth pair; and he will give up more roast beef for the fourth pair of shoes when he is consuming 8 pounds of roast beef a month than when he is consuming 3 pounds a month. The additional money he would spend to get a fourth pair of shoes is the worth to him of the fourth pair of shoes; it is the *marginal worth* of 4 pairs of shoes. If the *marginal worth* to him of 4 pairs of shoes is greater than the price of the shoes he will buy the fourth pair; if it is less than this he will stop at 3 pairs and spend the remaining dollars on the other alternatives.

Thus the way in which John spends his money provides an index to his preferences. He will tend, however roughly, to adjust his expenditures in such a way that he is "getting the most for his money." If spending one dollar less on shoes and one dollar more on socks would increase his satisfactions, he will change his spending pattern in this way. Given his income, his tastes, and the relative prices of the different commodities he buys, John will tend to distribute his spending so that each dollar does as much as possible, and no diversion of a dollar from one kind of spending to another would bring any better results. He will add to his per annum purchases of shoes so long as another pair adds more to his satisfactions than the price of the shoes is worth to him in other things.

This leads us to an important conclusion: *The price of a commodity is a rough measure of the marginal worth of that commodity to each of the individual consumers buying it.* Suppose the price of a serving of roast beef to be \$.20. To Tom, who at this price eats 10 servings of roast beef a month, the worth of the tenth serving is at least \$.20; the tenth serving of beef is at least as important to Tom as is any other thing on which he might spend the \$.20. To John, who eats 8 servings a month, the worth of the eighth serving is at least \$.20; the eighth serving of beef is at least as important to John as is any other thing on which he might spend the \$.20. To Dick, who eats only 3 servings of roast beef a month, the worth of the third serving is at least \$.20; the third serving of beef is at least as important to Dick as is any other thing on which he might spend the \$.20.

We may now take the final step in this discussion of indexes of individual preferences. If the price of shoes is a measure of the marginal worth of shoes to each of the consumers of shoes; and if the price of ties is a measure of the marginal worth of ties to each of the consumers of ties; and if the price of roast beef is a measure of the marginal worth of beef to each of the consumers of beef; and so on; then we may say that the marginal worth to each consumer of a product that sells for \$8 is greater than the marginal worth of another product which sells for \$7, or for \$6.

If shoes cost \$8 a pair and ties cost \$2 each, and if John buys 3 pairs of shoes and 15 ties, then it must be that the third pair of shoes is more important to John than is the sixteenth tie. Otherwise John would buy only 2 pairs of shoes, spending \$2 for another tie and he would still have \$6 left over for other things. It is a safe bet that if the prices of shoes and ties were the same he would change his consumption in favor of shoes. *When the price of one commodity is higher than the price of another commodity, there is a strong presumption that an additional unit of the first commodity would be worth more to each consumer than he would lose in giving up one unit of the second.* We are thus provided in the market with a rough measure of the relative worth to each consumer of one unit more or less of some products as against one more or less of other products.

Consumer Preferences and the Revenues and Costs of Firms

Consumer spending is the source of the revenues of sellers; it is a direct source in the case of retailers, an indirect source in the case of firms from which retailers buy, and so on back through the many stages of productive activity. When they express their preferences through spending, consumers thus transmit back to entrepreneurs votes in the form of revenue opportunities that provide the incentives to productive activity. Entrepreneurs are not primarily interested in getting resources allocated in accordance with consumer interests; they are primarily interested in maximizing their incomes. But it is from consumers that these incomes ultimately must be derived. We have already indicated in some detail in Chapter 12 that the interests of producers and consumers are sometimes in harmony, while sometimes they conflict. We are now ready to examine more fully the implications of that analysis for the purely competitive case.

The sale values of marginal products as indexes of consumer preferences

Let us suppose that we wanted to compare the value to consumers of increasing the output of silk scarfs or of playing cards; and that silk scarfs sell at \$1.00 each while playing cards sell at \$.50 a pack. If it were a question of adding one silk scarf or one pack of playing cards, and if the addition to costs involved were the same in the two cases, the answer would be quite clear: we would rather have the scarf. But the question is not quite as simple as that. Suppose that adding another unit of variable agents in order to expand a firm in the silk-scarf industry would increase the output of scarfs by 100, while adding another unit of variable agents in order to expand output in a firm in the playing-card industry would increase the output of playing cards by 300. What might we then conclude? The extra 100 scarfs would probably be sold to a large number of consumers (perhaps 100 of them), and to each of these individuals the marginal worth of another scarf would be approximately the \$1.00 which he pays for it. The extra packs of playing cards would similarly be sold

to a number of consumers to whom the worth of the new pack would be approximately \$.50. The sale value of the marginal product of 100 scarfs is 100 times \$1.00, or \$100; this is a rough index of the aggregate worth to consumers of the additional scarfs. The sale value of the marginal product of 300 playing cards is 300 times \$.50, or \$150; this is a rough index of the aggregate worth to consumers of the additional playing cards. What may we now conclude? Apparently the marginal product of 300 playing cards is preferable, from the consumer point of view, to the marginal product of 100 scarfs. The decision as to which would be the better thing to produce will depend on what has to be given up to expand output in each of the two products, and if the sacrifices are equal, the decision will clearly lie in favor of the playing cards. Since the sale value of the marginal product in any firm is a measure of the worth to consumers of that product, it is often designated as the *economic contribution* of the additional unit of input involved in making that marginal product available.

Opportunity costs and sale values of marginal products

In a purely competitive economy the costs involved in attracting the last worker or ton of coal or bale of cotton or entrepreneurial labor and investment into a firm in the cotton textile industry would provide a measure of the worth to consumers of the services of that worker or ton of coal or bale of cotton or entrepreneurial labor and investment in alternative employments. The costs of attracting agents are the amounts anticipated of which is necessary to persuade them away from alternative employments.²

These costs are called *opportunity costs* because the reward necessary to attract an agent is (ignoring motives other than the maximizing of incomes) the amount the owner of the agent would get in his best alternative opportunity. Clearly the producer of coal will sell where he can get the highest return, the seller of his own labor services will take the job where pay is highest. In order to get another ton of coal or another laborer an entrepreneur must pay for them at least as much as would anyone else. The amount the owner of this agent would receive in the best opportunity that is foregone is the "opportunity cost" of attracting the services of the agent.

Under conditions of pure competition, opportunity costs of marginal agents in each firm are therefore a measure of the worth to consumers of the other things these marginal agents might otherwise produce. This statement is of basic importance. Some further explanation of its meaning is therefore necessary.

1. Under conditions of pure competition, entrepreneurs in the cotton cloth industry will seek to employ additional units of input (labor, coal, and so on) so long as average expenditure (input price) is less than the sale value of the marginal product that would result. (This follows from

² Assuming supply of each agent to the industry is not perfectly inelastic.

the fact that under conditions of pure competition marginal revenue equals the sale value of the marginal product, and marginal cost equals average expenditure.)

2. The price that will have to be paid to attract the thousandth spinner into any firm in the cotton cloth industry, or to attract the ten-thousandth ton of coal, will in each case depend on the alternatives open in other industries to that worker or seller of coal. It will be equal to what can be obtained by the *marginal* agent in its best foregone opportunity—for example, by the worker who is just “on the fence” in deciding to accept this opening in the textile firm rather than that one in shoe manufacturing. From this point of view, the *marginal* worker is the man who is just “on the margin” of coming into or dropping out of (any firm in) an industry.

3. But entrepreneurs in the shoe industry will bid to get and retain workers up to a wage that equals the sale values of the marginal products of the workers who just come in or stay in that industry. The wages of workers in the shoe industry (and the price of a ton of coal in the shoe industry) will thus tend to equal the sale value of the marginal product of a worker (or a ton of coal) in that area of production. A cotton cloth manufacturer will have to pay to get another worker at least as much as the wages available to him in shoemaking; *therefore he will have to pay a wage equal to (but not more than) the sale value of the marginal product of a marginal man's labor in shoemaking.*

4. We have already seen (page 274) that the sale value of the marginal product of agents employed in one firm may be compared with the sale value of the marginal product of these agents in another firm to determine the relative importance to consumers of the two alternatives. We now see that if pure competition is assumed, the wage that must be paid to attract marginal units of productive agents from their best alternative employments is an amount equal to the sale values of their marginal products in these alternatives. *What is average expenditure to a purely competitive buyer of the services of an agent (say coal in cotton cloth manufacturing) may therefore serve as an index of the worth to consumers of the services of a marginal unit of that agent (coal) in its best alternative use.*

In this analysis we have taken our illustrations from cases in which productive agents are hired and costs are therefore explicit. The same conclusions apply, however, when costs involved are merely implicit. The opportunity costs of attracting the investments and labor services of entrepreneurs into an industry reflect the worth to consumers of the contributions of these entrepreneurs in their best alternatives.³ All costs

³ Suppose, for example, that there are one hundred men in a certain area who become entrepreneur-tailors. The illustration may be simplified further by assuming that they are all equally successful tailors. For some of these men the best alternative to tailoring is very good, for others it is very poor. If returns from tailoring were to drop considerably, those tailors with good alternative opportunities would gradually shift out of tailoring. The entire one hundred men would go into tailoring and remain in it only if the implicit expected return resulting from

(both explicit and implicit) are "opportunity costs": they measure the worth to consumers of the services of a marginal agent in its best alternative employment.

Conclusion: The Allocation of Resources Under Pure Competition

Actions intended to maximize incomes (including profits) would tend to bring about an allocation of resources that approximated very closely a realization of consumer preferences according to individual rankings in their dollar influences on economic activity. Preceding analysis has already provided sufficient proof of this statement; it would be true in both short- and long-run adjustments. We may now draw the thread of discussion together in a recapitulation.

Short-run allocation of resources under pure competition

In the short run only a part of the productive agents can be shifted from one firm to another, from one industry to another. These agents that are "variable" in the short run may be employed in greater numbers in one capacity or in another. In the short run under conditions of pure competition a seller will add units of input (and hence increase outputs) so long as to do so adds less to cost than it adds to revenue. The amount added to the firm's revenue is under pure competition the same as the sale value of the marginal product. The sale value of the marginal product indicates the worth to consumers of the additional output that will result from employment of another unit of input. Costs depend on the amounts that productive agents could obtain in foregone opportunities, and these "opportunity costs" reflect the amounts consumers would be willing to pay for the marginal product of these agents in these "other" employments. Therefore, when under purely competitive conditions it adds more to income than to costs to add a unit to input, this means that consumers are willing to pay more for the marginal product of these agents employed here than for what they would add to output in other firms and hence in other industries. If on the other hand the addition of a unit to input would add less to revenues than to cost it would indicate a preference on the part of consumers for the marginal product of these agents employed

their tailoring activities was sufficient to match the foregone opportunity of the tailor with the best alternative. These alternatives represent opportunities to make incomes by making economic contributions to consumers in other ways. If for a certain period actual incomes of entrepreneurs in tailoring more than cover the opportunity costs of attracting the labor and funds of the most reluctant tailor (the marginal offerings of tailoring services), the excess is a profit. More men will be attracted from yet better alternatives into tailoring until returns here are just equal to the opportunity costs of the now marginal tailor. If actual incomes of the hundred tailors were less than the alternative opportunity for the marginal tailor or tailors among them, however, these marginal men would shift out of tailoring into their now better alternatives. In the equilibrium adjustment tailors would receive incomes just equal to those available to the marginal tailor or tailors elsewhere.

elsewhere, and sellers would find it advantageous to respond to these consumer preferences. *Within the range of possible adjustments open in the short run, producers are responding consistently with consumer wishes as expressed in the market.* Variable agents are allocated consistently with consumer preferences within the short-run situation.

Long-run allocation of resources under pure competition

An examination of long-run competitive equilibrium shows that when the maximization of profits is the entrepreneurs' goal and agents of production are free to move into all fields, all these agents will tend to find employment where they are most wanted by consumers because it is here that they will receive the greatest reward.

Let us examine this proposition further. Suppose that producers in a given industry are losing money. Since costs of the agents they employ (and their own salary allowances) are determined by the sale value of the marginal products that would result from their employment in alternative uses, the fact that an industry is losing money can mean only one thing. Some of the agents (fixed and variable) employed in this industry would in alternative employments turn out products worth more. The greater value of their product in alternative employments arises out of the willingness of consumers to pay more for what they would turn out there than for what they would turn out in the losing industry. Resources are not being allocated in accordance with consumer preferences. Can this situation persist? Obviously it cannot. Agents of production will leave the losing industry and find employment where they will be paid more.

Now let us suppose the opposite situation. Suppose that consumers are willing to pay high enough prices for the product of an industry so that agents in this industry are earning more than in alternative employments. Producers in this industry are making profits. Resources are again maldistributed from the consumer point of view, since consumers are willing to pay more for them here than elsewhere. In the long run under pure competition new firms will appear, more productive agents will be employed in the profitable industry and there will be an increase of output in that industry. *Thus we see that under conditions of pure competition there will be a tendency for agents to move into an equilibrium adjustment in which the behavior of producers will be entirely consistent with the preferences of consumers.* So long as they have not attained this position, there will be an advantage to some producers in making shifts. On the other hand, there will be no incentive to move away from this position, since any such shift involves a decrease in income to those making the change. Any such shift would lead to the employment of agents in a use desired less by consumers, that is in one for which consumers are willing to pay less.⁴

⁴ The most important flaws in purely competitive adjustments from the consumers' point of view center in discrepancies between social and private costs and gains.

Social costs may be greater than private costs; that is, there may be hidden costs

that are not paid by the producer of a good but by third parties who happen to be damaged. These costs are not included in the producer's cost calculations and are therefore inoperative as influences helping to determine equilibrium adjustments. Price will be lower in such an industry and output will be greater than the true costs justify. An example of this situation would be the depreciation of property values resulting from smoke and gas escaping from nearby factories. This property depreciation is a part of the social cost of manufacture of the article produced in the offending factory, but is not included in the manufacturer's cost calculations. Social regulations are sometimes imposed in such cases, the requirement of smoke consumers in some cities being a case in point. Occasionally special taxes are levied on offending industries.

Similarly there may be hidden gains accruing to third parties who make no payment for them. These gains are excluded from the price paid the producer. They are returns (unpaid) on a by-product of his enterprise. An example of this situation would be the appreciation (increase) of property values downstream as an indirect result of flood control incidentally provided by private power dams. These property owners make no payment for the service they receive. The amounts they would be willing to pay are therefore omitted from the revenue estimates of the dam builder, since his estimates are based only on the prices he receives for goods produced by the use of his power dam and offered for sale.

Part IV

MONOPOLISTIC AND MONOPSONISTIC MARKETS

PREAMBLE

A LARGE proportion of markets are monopolistic: an entrepreneur can raise the price of what he sells by limiting the quantity offered for sale. Also, many markets are monopsonistic: an entrepreneur can lower the price of what he buys by taking a smaller quantity. People like to get high prices for what they sell and pay low prices for what they buy. Entrepreneurs can increase their profits by pushing prices around, except that they cannot charge prices "too high" without losing "too many" sales, and they cannot pay prices "too low" or they will not be able to get "enough" of what they want to buy. What they will do about prices, and about the level of production, will depend on a number of things—partly, for example, on whether there are a lot of producers of close substitutes, or whether substitutes are very remote; partly on whether there are just a few rivals, who react and retaliate to policies initiated by any one among them. It is now time that we look into these things.

Moreover, businessmen are very inventive in their search for profits, and they do not confine their inventiveness to making new machines and materials. They try all sorts of ways of winning customers. In part they use price cutting for this purpose. But they also vary the characteristics and qualities of the goods they produce for sale; they try to make things that will appeal to consumers and will be purchased readily. They offer convenience of location, easy credit terms, attractive shopping surroundings. They evolve elaborate techniques for persuading consumers to buy their products, through store displays, magazine advertising, and a hundred other practices. Sometimes they give information and sometimes they give misinformation. When competition for customers by lowering prices is "hamstrung" by formal agreements or the development of customary prices in an industry, entrepreneurs are especially encouraged to win trade by these other means. The effects of these practices are extremely important to both businessmen and the general public that buys their wares.

CHAPTER 21

Simple Monopoly and Simple Monopsony

WHEN a firm is the sole producer of a product, and when rivals producing substitutes are so remote that the entrepreneur of this firm can reasonably ignore any reactions of rivals to his sales policy, the situation has been designated "simple monopoly." Analogously, when a firm is the sole buyer of a productive service, and when rival outlets are so remote that the entrepreneur of this firm can reasonably ignore any reactions of rivals to his buying policy, the situation has been designated "simple monopsony." We have already some acquaintance with these types of situations, but we are now ready to examine them in greater detail.

Simple Monopoly

Exclusion of new firms as the basic foundation for simple monopoly

Exclusive control by one firm over the production of a product which has no close substitutes may come about in a number of ways. Moreover, when one firm controls as much as, say, 90 per cent of the production of such a good and is able to retain this position in the market, it approximates very closely the characteristics of a simple monopoly.

The most clear-cut cases of simple monopoly arise out of situations in which there is a definite method by which the entry of any other firms may be prevented. This may occur if a firm holds a basic patent either on the end product or on the methods of manufacture that effectively keeps out other concerns. It may occur, as in the case of aluminum up to the war, through sole ownership by a firm of virtually all the natural resources that are essential in the manufacture of a product.

The development of techniques of business organization that make possible very large combinations in which many corporations are joined is a fundamental basis on which may ultimately be built a dominant position of a single firm in an industry. This dominance may become so great as to lead to what is essentially simple monopoly, especially when the large organization bulwarks itself through manipulations of railroads or banking facilities, or through other direct and indirect unfair competitive tactics designed to eliminate new firms before they get well established and to scare away others considering the possibility of entering the industry.

Relation between "decreasing costs" and simple monopoly

When an increase in the output of an industry is accompanied by a decrease in the total unit cost at which production is possible, the industry is described as a "decreasing cost" industry. Decreasing costs arise only when the optimum scale of enterprise is so great relative to the demand for the product that the market cannot support more than one firm at that scale of enterprise at which the least-cost-combination is minimized.¹ Essentially, costs are "decreasing" because of the economies of scale of enterprise that would ensue if this were to continue to be a *one-firm industry*.

The nature of a decreasing cost industry may be made more evident by the use of graphic representation, as in Figure 21—1. Curves TUC_1 ,

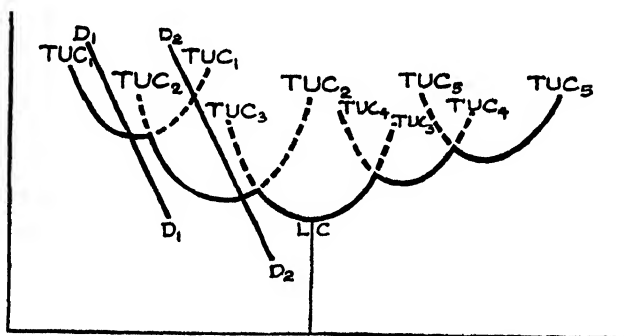


Fig. 21—1. Decreasing costs.

TUC_2 , TUC_3 , and so on, are total unit cost curves for different scales of enterprise in a single firm, say an electric power company. Each curve is drawn in a heavy line through the range in which it is lower than any other, and is extended in dotted lines when it rises above other total unit cost curves. The heavy scalloped line is therefore a curve that indicates the total unit cost at which each output can be made available. Through the first three cases this curve is generally falling; this is a condition of "decreasing costs." If demand is small relative to the optimum scale of enterprise (TUC_3), or if optimum scale of enterprise is large relative to demand, then the industry will be characterized by decreasing costs. In Figure 21—1 this would be the case if demand were as represented by DD_1 or DD_2 . Any expansion of output of the firm to the point LC (and hence, any expansion of the output of the industry, if this is a one-firm

¹ We are ignoring the "external economies" that may arise in a particular industry that in expanding is able to buy more cheaply than before from another industry that is a decreasing cost industry of the type we describe; we are also ignoring the organizational economies involved in very great expansions of an industry that change it from one in which there are many small firms (e.g. custom-made airplanes) to one of large mass production enterprises. (See the Appendix to Joan Robinson's *Economics of Imperfect Competition*.)

industry) will make possible lower costs per unit produced. Through this range this is a "decreasing cost" industry *because the economies of scale of enterprise within a single firm have not yet been fully exploited.*

Under what circumstances will such an industry remain a one-firm and hence a simple monopoly industry? Under what circumstances will a second firm appear, or perhaps a third, thus converting it into an industry characterized by duopoly or oligopoly? The first firm in a decreasing cost industry would of course much prefer that no other firm appeared to compete with it; but if profits are great, a new firm is likely to appear unless other restrictions on its entry are used. In a decreasing cost industry it would be quite impossible for two or more firms to cover costs if they competed actively with each other; demand for the product is not sufficient to support both firms at optimum scale, and active price cutting would eventually result in the expansion of the output of one firm and the elimination of the other from business. But instead of competing aggressively the two firms may enter into an agreement, tacit or formal, whereby price is maintained and each firm restricts its output. Both may then make some profits. This is obviously wasteful of resources if we look at the picture from the point of view of the general public; the two firms together are producing at much higher cost per unit of output than would be possible if the same aggregate output were produced by one firm only. But the entrepreneurs of such enterprises are interested primarily in profits, not in general social welfare. Thus decreasing cost industries do not necessarily remain one-firm industries. Only where it would be unprofitable for a second firm to enter (even anticipating a price-output agreement with the initial firm), will the industry remain in the hands of a single enterprise; only then will it be a simple monopoly merely because of the relation between demand and optimum scale of enterprise and in the absence of other restrictions on the entry of new firms.

Short-run maximum profit pricing of products of a monopoly

The average revenue schedule of a simple monopolist is the demand for the product of an industry, since the one firm is the industry. Substitutes are so remote that no one of them involves close interrelations with the monopolized industry, either in consumer demands or in the responses of rival producers to these demands. The "partial equilibrium" analysis of input-output adjustments of monopolistic enterprises as it was presented in Chapters 10, 12 and 13, therefore gives us a good first approximation to the analyzing of adjustments not only within a simple monopoly firm but, since the firm and the industry are the same, in the monopoly industry as well.

In Chapter 12 we analyzed input-output adjustments, arriving at the conclusion that monopolistic enterprises would curtail inputs and outputs according to the relation between marginal revenues and marginal costs instead of expanding by comparing the sale value of marginal products with marginal costs. This provides us automatically with the answer to

monopoly pricing. Given the consumer demand for a product, the determination of the output or quantity sold in an industry automatically tells us the associated price; price will simply be the amount consumers are willing to pay for that quantity of goods. Since under simple monopoly there is but one firm in an industry, and the demand for the product is the average revenue schedule of the firm, once we have determined the maximum profit output of the firm we may read from the average revenue schedule the price at which the product will be sold. *This will be a higher price than would be charged if the concern were to expand its output so long as the sale value of the marginal product exceeded marginal cost;* in fact it is just for the purpose of pushing up price that the monopolist thus restricts inputs and outputs. He is limited in his power to raise price by the fact that at higher prices more and more consumers withdraw from the market. The less consumers are discouraged from buying as price is raised, the higher will the maximum profit price be; in other words, the more inelastic the demand, the stronger the position of the monopolist in raising prices. But eventually there will always be a price above which demand is so elastic that it will not pay the monopolist to incur the sacrifice of sales involved in raising his price any higher.

The fact that a firm usually employs a number of productive services that it can use in various proportions even in the short run, and the fact that it frequently sells more than one product, does not change these general conclusions. The analysis of equilibrium of the firm under conditions of pure competition provides us with a method for handling this more complex and realistic set of conditions in the case of monopoly as well. If a monopolist were completely informed and rational in his consideration of short-run possibilities and if he ignored long-run conditions, he would arrive at the following adjustment: (1) Each variable agent would be employed up to a point where the addition of yet another unit of input would result in a marginal cost greater than marginal revenue, while curtailment of one unit of input would involve a reduction of revenue (marginal revenue) greater than the reduction of cost (marginal cost); *this implies inputs less than those that would follow from a comparison of the sale value of marginal product with marginal cost.* (2) Each of the joint products of the concern would be produced up to the point at which an expansion in output would add more to costs than revenues and a reduction would curtail revenues more than costs; *this implies outputs less than those that would follow from a comparison of the sale values of marginal products with marginal costs.* (3) There would be no other method of production or combination of products that would improve the profit position of the firm. The respective prices of the monopolized products of the enterprise would depend on the relation between demands for the products and the variations in costs that are associated with varying the input-output combinations of the firm; these things are taken into account in arriving at the maximum profit adjustment in input and output, and this adjustment involves a simultaneous determination of prices of products.

The effects of long-run considerations on the price policies of a monopolist

In deciding on current price policy, the entrepreneur of a monopoly concern would take into account some long-run considerations that do not show up in the immediate possibilities of the market. These long-run considerations would generally lead to a somewhat more moderate price policy than would be followed by a firm taking into account short-run factors only. Important among these long-run considerations are: (1) long-run versus short-run consumer demand, (2) potential competition from new firms, and (3) public attitudes and legal restraints.

1. Long-run consumer demand is likely to be more elastic than short-run demand, so that the adjustment of a monopolist to long-run considerations would lead him to modify his policy somewhat toward greater outputs and lower prices than if he based policy on short-run considerations only; it will not, however, lead him to go so far as to compare the sale value of marginal products with marginal costs. This greater elasticity of long-run consumer demand will affect the producer of consumer goods directly; it will be reflected back in a more elastic demand for the products of enterprises that contribute raw materials and equipment for the production of consumer goods, as well. The reason for greater elasticity of long-run consumer demand can be quite easily understood by considering how consumers would respond over extended time periods to relatively low or high prices. If prices are held at high levels, consumers may get used to going without the good, and consumption may be gradually curtailed; as time goes on this may spread to more consumers and each individual consumer may curtail more and more. On the other hand, the gradual adjustment of consumption habits to lower prices will have exactly the opposite effect; consumers will get used to the product, more people will be interested in it, and those already consuming it may increase their consumption if prices remain low over extended periods. Thus in the long run consumer responsiveness to price change may be, and usually will be, much greater than in the short run. The greater the difference between long-run and short-run demand elasticity, the more important will be moderation of short-run price policy in the interests of long-run profits.

2. Whether or not a concern seeking to maximize profits would reasonably take into consideration potential competition from new firms will depend on the kinds and strength of the fences it has built to keep new firms out of the industry. There may also be a question of balancing the costs of keeping the fences strong despite high price incentives or of modifying the prices and getting along with less expenditure on policies of exclusion. If a firm is very well established in a bulwark of privilege and exclusive position, owning indispensable patents, having exclusive control over essential raw materials, functioning in a closely integrated web of controls over banking and (especially in earlier times) transpor-

tation facilities, then it may pursue extremely high price policies without much concern for the competition that these prices may attract. If, on the other hand, the fences are weak, depending primarily on costly tactics of unfair competition and uncertain manipulation, and seriously threatened by potentially competing products and methods of production, then the fear of potential competition that may become real, if price policy is extreme, will become an important consideration in the action of the monopolist entrepreneur.

3. One of the most real considerations of a monopolist deciding on price policy is the state of public opinion. Public hostility to what is considered unfair practices and exploitation appears in many concrete forms. Those most feared by producers are probably consumer boycotts, both formal and informal, and the clamping down of legal restrictions and controls. The monopolist may find that a rise in price is met by consumer antagonism so great that potential buyers refrain from taking goods that they would in fact like to have at the price asked. They do not buy the product, because they consider the price "unfair." Moreover, the monopolist may find that his flight into monopoly exploitation is cut short by the activities of legislatures, the Federal Trade Commission, or other government bodies guarding social interests. Expressed in either form, hostile public opinion is too important to be ignored, and may in many instances temper the behavior of monopolists seeking to maximize profits.

Errors in estimating maximum profit positions

It is not realistic to assume that monopolists always act with omniscience and extreme care in the effort to arrive at the maximum profit point. The combination of inertia, complexity of the problem, lack of knowledge as to both short-run and long-run demand conditions, and uncertainties as to the seriousness of potential competition and legal restraints, may result in the charging of prices either greater or less than would maximize profits in either short or long runs. Although this argument is sometimes presented as a reason why less than the maximum profit price would be charged, there seems to be just as much reason for assuming that it would lead monopolists to set prices higher than that point. In many cases the extent of possible increase in sales by setting lower prices is unappreciated. When a monopolist assumes the demands for his products to be less elastic than they are, he sets his prices higher than those that would yield him maximum profits, and he sells smaller outputs.

Scale of enterprise

Many factors affect the scale of enterprise of a simple monopoly. Some of these factors are common to competitive situations as well. The considerations peculiar to a monopoly may, however, be grouped under two main headings: (1) Restrictions of scale in order to maximize profits in the absence of threats from potential competitors, and (2) expansion of

scale of enterprise as a result of tactics designed to gain and maintain a position of dominance, if possible an exclusive one.

1. In the long run an entrepreneur takes into consideration variations in the use of all productive agents, including those that in the short run are "fixed." Therefore the curtailment of inputs that applies to only variable agents in the short run will enter into the entire production plan in the long run as a part of the technique for maintaining high prices. This was strikingly illustrated in the underdevelopment of aluminum plants in America prior to the war.

2. On the other hand, very uneconomic expansion of plant capacity sometimes occurs as an incidental result of tactics designed to maintain exclusive control in an industry. An extreme illustration is the early history of the Standard Oil Company. Competing concerns were bought outright, their plants thus acquired becoming part of the equipment of the big concern fast approaching a position of complete monopoly. Some of these plants were shut down; others were operated. In most cases it was not any greater efficiency of large-scale production that dictated their purchase but merely intent to attain a monopoly position. Such tactics may even lead to the appearance of new firms started with the intent and for the purpose of selling out. Overzealousness in building monopoly-control through such means has sometimes turned out to more than neutralize the gains from the ensuing monopoly prices.

Profits in the long run

One of the characteristics of simple monopolies that have led to considerable public hostility against them is their ability to obtain profits over extended periods of time by holding up prices of their products. Where restrictions on the entry of new firms are very strong large profits may persist almost indefinitely, though almost any monopoly is ultimately threatened by the development of new processes and new products even barring all other incursions into its markets.

Simple monopoly in the most exact sense means the existence of one firm only. However, one very large dominant concern may suffer the continued existence of a few small enterprises that follow its lead and that take only a very small portion of its market. The concern is still essentially a monopoly. We might regard it as a hen with a few very tiny baby chicks. Such a situation differs from the poultry-house in that the baby chicks usually remain on such restricted diets that they fail to grow up; but it resembles the poultry-house in that there are potential competing hens among the chickens. If profits are very great the situation is clearly unstable and eventually one or more of the small concerns is likely to grow up to be a real threat to the monopoly control of the dominant firm. Meanwhile it may, however, be held back for a long time by repressive strategy of the dominant enterprise.

During this time monopoly profits are partially checked both because of the costliness of the tactics engaged in to keep competitors out or to

keep them from growing and because of the nibbling of these competitors along the way. But even where the exclusive position of the dominant concern is thus unsecure, profits will persist far longer and in far more extreme form than in purely competitive situations. Moreover, the appearance of competitors results not in pure competition but in oligopoly, which may build up restraints by the group of firms on the further entry of outsiders. The result may be the persistence of profits in the now oligopoly situation.

Simple Monopsony

The occurrence of simple monopsony

Simple monopsony in the buying of productive services is sometimes associated with simple monopoly in the sale of the product. Where productive services are specialized to an industry and that industry is in the hands of one firm only, the single firm is then both a monopolist in the sale of the product and a monopsonist in the purchase of the specialized productive services. Approximations to these situations may be cited, but full-fledged examples are hard to find. This is true, in part, because the very basis of the monopoly is sometimes the ownership of productive resources specialized to the industry, and therefore there is no exchange between the firm and someone from whom it buys the services of these specialized resources. It is true, in part, because many of the productive services used in the monopoly industry have alternative outlets in other industries as well.

The other most common basis for an approximation to simple monopsony is in geographic barriers between firms. These barriers may be relatively unimportant in the sale of the product of the enterprises involved, and yet very important in the buying of productive services. Some of the most commonly cited examples are the geographically separated markets for labor; but equally important are some instances in which monopsony control is approximated in the purchasing of agricultural products in different regions. Where immobility is extreme the competition of firms buying in other geographic markets may be very remote; where the productive agents and their services are highly mobile, however, buyers in different places may compete closely for these services and the market is then characterized by monopsonistic competition.

Short-run maximum-profit pricing of productive services by a monopsonist

The average expenditure schedule of a monopsonist is the entire supply of this particular productive service, since the monopsonist firm is the sole outlet for this service. Rival buyers of the services of the productive agents involved are so remote that no one of them plays a significant role in the situation. The "partial equilibrium" analysis of input-output adjustments of monopsonistic enterprises as it was presented in Chapter 12 therefore gives us a good first approximation to the analyzing of ad-

justments not only within the firm but, since the firm is the sole market for the particular productive service, in the entire market as well.

In that analysis we arrived at the conclusion that monopsonistic enterprises would curtail inputs according to the relation between marginal cost and marginal revenue instead of expanding by comparing average expenditure and marginal revenue. This provides the answer to monopsony pricing. Given the supply of a good or service, the determination of the quantity taken automatically tells us the associated price; price will simply be the amount sellers require if this quantity is to be forthcoming. Since under simple monopsony there is but one firm buying a productive service, and the supply of that service is therefore the average expenditure schedule of the firm, once we have determined the maximum profit input of the firm we may read from the average expenditure schedule the price at which the service will be purchased. *This will be a lower price than would be paid if the concern were to expand its input so long as average expenditure was less than marginal revenue*; in fact it is just for the purpose of pushing down price that the monopsonist thus restricts inputs. ~~He is limited in his power to lower price by the fact that at lower prices more and more productive services will be withdrawn from the market.~~ The less sellers of these services are discouraged from selling by lower prices, the lower will the most profitable input price be; in other words, the more inelastic the supply of the service, the stronger the position of the monopsonist in pushing down the price. But again, eventually there will always be a price below which supply is so elastic that it will not pay the monopsonist to incur the sacrifice of production involved in pushing the price of a productive service down any lower.

The facts that firms vary the proportions of productive services used and that they turn out several products instead of one are again no causes for concern in applying the principal conclusions of partial equilibrium analysis. The problem is essentially the same as in the generalization of any partial equilibrium analysis of adjustments of the firm. The monopsonist, if he were completely informed and rational in his consideration of short-run possibilities, and if he ignored long-run considerations, would arrive at the following adjustment: (1) Each variable agent would be employed up to a point where the addition of another unit of input would result in a marginal cost greater than marginal revenue, while the curtailment of one unit of input would involve a reduction of revenue (marginal revenue) greater than the reduction of cost (marginal cost); *this implies inputs less than those that would follow from a comparison of average expenditure with marginal revenue*. (2) Each of the joint products of the concern would be produced up to the point at which an expansion in output would add more to costs than to revenues and a reduction would curtail revenues more than costs; *this implies outputs less than would follow from a comparison of average expenditure with marginal revenue*. (3) There would be no other method of production or combination of products that would improve the profit position of the firm. The respective prices of the monopsonized productive serv-

ices used by the firm would depend on the relation between supplies of the different services and the variations in revenues associated with varying input-output combinations of the firm; these things are taken into account in arriving at the maximum profit adjustment in input and output, and this adjustment involves a simultaneous determination of prices of the productive services used.

The effects of long-run considerations on the price policies of a monopsonist

Just as a monopolist selling his product or products will take into account some long-run considerations in determining price policies, so also will a monopsonist consider long-run factors in arriving at decisions concerning the pricing of productive services. And here again these long-run considerations would generally lead to a somewhat more moderate price policy than would be followed by a firm taking into account short-run factors only. Important among these long-run considerations are: (1) long-run versus short-run supply of the productive services, (2) potential competition from new firms, and (3) public attitudes and legal restraints.

1. The short-run supply of a productive service is likely to be much more inelastic than is the long-run supply of the same service. This is due to the greater mobility of productive agents in the long run. If the agents are laborers, offering their labor services for sale, they may eventually move from one place to another even quite remote if the opportunities there are significantly better. Similarly workers may learn new skills and desert old ones that have ceased to be lucrative; and more important, as older workers drop out of the labor markets they will not be replaced by younger men in the same skills unless these skills bring in incomes that appear reasonably attractive. If the agents are machines or raw materials, they will also eventually cease to be replaced in present form if the payments for their use are too low, even though in the short run they may be obtainable at very low prices or rents. Insofar as long-run responsiveness to the depressing of the price of a productive service evokes greater withdrawals of that service than would occur in the short run, a consideration of long-run reactions will lead the monopsonist firm to pursue a more moderate policy in the payment for the service; he will exert some downward pressure on price, but a pressure much less than would maximize his short-run profits.

2. If a monopsonist firm is in an exceptionally favored position in the buying of productive services, so that monopsony profits are being obtained, other entrepreneurs will of course seek to share this opportunity. If barriers to the entry of other firms into these buying markets are strong, the monopsonist may remain unchallenged in his comfortable position; this situation is most likely to be approximated when the monopsonist is also a monopolist. On the other hand, if barriers are weak or the maintaining of them is a costly proposition, the firm may conceivably hesitate to exploit its full possibilities. It may be that potential competi-

tors can be kept out fairly easily if the incentive to intrusion is not too great; but that extremely low prices for productive services will lead to the cracking of the barriers. Under such circumstances, a monopsonist would modify his short-run low-price policy in view of long-run potential competition from new buyers.

3. Public attitudes and legal restraints are important in checking the actions of monopsonists as well as those of monopolists. They take somewhat different forms in most cases, however, and are more closely allied to strong sentiments of humanitarianism. Where monopsony is strong the victims of the downward pressures are frequently less able to cope with the situation than are consumers of monopoly-produced goods, since the entire livelihood of an individual or family may come from the selling of the labor or the potatoes or the tobacco the price of which is being depressed. Boycotts of particular goods by consumers are far less drastic in their effects on these consumers than are strikes of laborers who must go without incomes during the strike, or of small farmers whose sole livelihood is from their crops.

CHAPTER 22

Monopolistic Competition and Monopsonistic Competition

MONOPOLISTIC competition and monopsonistic competition are really hybrid types of market situations falling between simple monopoly and monopsony on the one hand and pure competition on the other. The characteristics of monopolistic competition are that there is a large number of firms selling products in approximately the same degree of rivalry with each other, and that each firm is producing a product slightly "differentiated" from the offerings of the others. The characteristics of monopsonistic competition are that there is a large number of firms buying productive services in approximately the same degree of rivalry with each other, and that each firm provides a slightly "differentiated" outlet for the services of the productive agents involved. Under these circumstances each firm has a slight degree of control over the prices of what it sells or what it buys, as the case may be; but the general range within which it exerts this control is narrow. That range is determined roughly, as in a purely competitive situation, by the mass effects of the actions of aggregates of individuals.

Monopolistic Competition

The bases of product differentiation

When variations appear in the form or in the conditions of sale of the same general product this situation is usually described as "product differentiation," and each variety is said to be a "differentiated product." Any condition that leads buyers to prefer one seller to another is a cause of "differentiation," whether this arises in variations of the form of the product itself or in the conditions and the services associated with its sale.

John may prefer Marlboros to any other brand of cigaret. This may be because he prefers the tobacco, considers the packing better, likes the texture of the tip. It may be that if he were given a blindfold test he would be unable to distinguish Marlboros from any other brand; but whether he could distinguish them or not, so long as he is under the illu-

sion that they are really different and to him superior he will be willing to pay more for a package or carton of Marlboros than for any other brand from the same counter at the same time. Marlboros are, so far as John is concerned, differentiated from other brands of cigarettes. The difference, whether real or illusory, may be built up in his mind by the recommendations of friends and the claims of advertisers or it may be based on his own experience and observation. The producer gains and retains his customers by advertising and the use of an identifying brand name quite as much as by price competition. If he provides something in his cigarettes which really distinguishes them from others and draws a special group of customers we may describe his action as "quality" competition in contrast to "price" competition.

It may be that Mary is entirely indifferent as to which brand of cigarettes, of breakfast food, or of canned grapefruit, she obtains, but there may still be product differentiation with respect to conditions of sale. She will pay more, if necessary, to get the cigarettes in the restaurant where she is eating or to get the canned fruit at the store around the corner from her house than she would have to pay to get the same goods elsewhere. Mary may prefer one cheese jar to another and will select the cheese on this basis, she may have a weakness for cellophane wrappings which leads her to prefer goods packaged in this form, and she may be ready to pay some premium to have the butter cut into quarter pound sections instead of getting the same butter in bulk out of one large container. The convenience of charging purchases and having them delivered may lead Mary to trade at the grocery or department stores where this service is available. She may have become friendly with her butcher and for that reason when the butcher moves to another store he may take her trade along with him; or she may value the comments of an especially cooperative salesman in buying shoes and always return to him. If Mary is short of time she will shop where she expects quicker service even though she has to pay more to get it. The atmosphere of the better clothing departments may be such a relief to her in contrast to basement crowds that she shops there even if and when she could get exactly the same good for less in the basement. She may value the prestige of a Bergdorf Goodman or a Saks Fifth Avenue label on her coat.

These small differences that lead us to prefer one good, one store, one salesman to another are very common. They are based in part on real differences in form and in services associated with sale; in part the differences are illusory. But whenever one seller is differentiated from others in the minds of consumers, he is able to charge a price slightly higher than some of the others and still retain some of his customers. The other sellers may also have customers especially attached to them and as a result some degree of control over price. If there are many sellers who are close rivals and yet differentiated one from another, each seller will have individually only an insignificant effect on other sellers in the general market for the product, at the same time that each seller can individually exert some price control over his particular variety.

Short-run price adjustments

In order to understand the pricing process in markets characterized by monopolistic competition it is necessary to look back at both the purely competitive and the simple monopoly types of conditions. The seller of a slightly differentiated product, competing with many rivals, has but a limited range within which he can control the price of the product he sells. Within this range he will maximize his profits by an input-output adjustment based on a comparison of marginal cost with marginal revenue; and marginal revenue will be slightly less than the sale value of the marginal product so that this adjustment will involve some curtailment of production. Since substitutes are very close, however, the discrepancy between marginal revenue and the sale value of the marginal product will usually be small; the average revenue schedule of the firm will be highly elastic, approaching the extreme case of perfectly elastic average revenue of a purely competitive seller.

The limits of the general range within which the single firm can control price will be determined by a rough relation between the aggregate offerings of all the rival firms in the general "production area" (for example all the producers of "ordinary" soaps) and the aggregate demand of consumers for the entire group of slightly differentiated products. If the consumer demand for soaps in general were to increase, each of the individual producers of slightly differentiated soaps would find that his average revenue schedule would be higher than before; and conversely, if the consumer demand for soaps were to decrease, his average revenue schedule would be lower. Similarly, if one of the major ingredients of most soaps were to become more costly, each of the individual producers of a particular variety of soap would tend to cut production somewhat and the aggregate effect would be a higher general price level for soaps.

Quality competition

There are three principal ways in which a seller may seek to attract customers away from his rivals: (1) by lowering price, (2) by changing the quality or form of what he offers for sale, including the conditions under which it is sold, and (3) by advertising.

The first is described as "price competition," and has been the focus of analysis in preceding chapters.

The second involves an appeal to customers by meeting their preferences in the form of goods and the conditions of sale, and is called "quality competition." Under pure competition, many sellers find themselves in a position to offer identical goods for sale at the prevailing price, and unable to improve their position in this respect by any changes open to them. In fact pure competition appears in a market situation in which many sellers acting freely to maximize profits tend to a position in which identical goods will be offered for sale at a uniform price. When each of many sellers differentiates his product and the conditions of sale in order to

attract customers, monopolistic competition appears. A seller will make special services available when he believes that to do so will be profitable, that is, if he believes consumers to want them enough to pay for any extra costs. He will eliminate such services—"cash and carry stores" do—when he believes that enough consumers would prefer to pay less and go without the special service. He will put more chromium trim on automobiles, more pleats in skirts, or feathers on hats, if he thinks that by doing so he can get enough more customers than otherwise, and/or charge enough higher prices to make these changes profitable. He will increase the durability of fabrics if he thinks consumers prefer this quality and are ready to pay the difference. The competition among sellers to get customers by offering preferred qualities results in pressures on all sellers to improve quality in response to consumer demands just as price competition exerts pressures on all sellers to lower prices and costs.

The third way in which a seller may seek to attract customers is by creating and exaggerating differences in their minds through a wide range of advertising techniques. Instead of merely responding to consumer preferences for one variety of good and one kind of service over another, the firm is then seeking to create the demand for which it produces. This aspect of competitive behavior will be examined in Chapter 24. For the present its existence will be ignored.

Long-run adjustments

In the long run what will be the adjustments in the profit positions of individual firms, in the number of firms in the "production area" of a whole range of varieties of one general product, in the prices at which these varieties will sell? Again, as in other types of market situations, new firms will appear wherever there are profit opportunities and wherever the entry of such new firms is not restricted. While under purely competitive conditions the entry of new firms involved the duplication of the products of existing firms, under conditions of monopolistic competition this means the appearance of new slightly differentiated varieties or brands, and hence somewhat different equipment. As new firms appear offering new varieties of the same general product they soak up part of the general market, taking some customers away from some of the previously established firms and reducing their profit opportunities. New firms will continue to appear so long as profits are available, thus tending to diminish the average earnings of established firms. A position of equilibrium will be reached when there are no more such opportunities unexploited, and when no firms remain in the production area operating at a loss. After all profit opportunities for new firms have been exploited, in some firms there will probably remain small profits but in most cases these will have been markedly checked. The range of prices of the differentiated varieties of the general product will reflect a rough equation between the demand of consumers for the entire group of goods and the offerings of the aggregate of slightly differentiated sellers.

Monopsonistic Competition

The basis of "outlet differentiation"

Any condition that leads sellers to prefer one outlet for their services to another is a cause of "outlet differentiation" analogous to the product differentiation resulting from preferences of buyers. It may be that the purchasing agents of one firm are more pleasant to deal with than those of another, and hence sellers have a slight preference for this outlet for their services. It may be that some buyers are more trustworthy or pay their bills more promptly than others and hence are regarded as more desirable outlets for those selling productive services. Such differentiation will appear whether the productive service being sold is the service of machines, of raw materials, of land, or of labor. But the most important and the most frequent causes of differentiation in outlets appears in connection with labor markets. The man selling a machine is not at all concerned as to how or where it will be used by the purchaser. The man selling his own labor is very much concerned with the conditions of its use, with the temperament of his "boss," the hours of work, the air and light, the region and the community in which the job is located, and many other matters. These preferences of workers are quite independent of the motive of maximizing money income, and share much in common with the factors that lead ultimate consumers to differentiate between one retailer and another.

Whenever one buyer is differentiated from others in the minds of the sellers of productive services, he may be able to pay a price slightly lower than some of the rival buyers and yet to obtain some productive services. The other buyers may also have sellers of productive services especially attached to them and as a result some degree of control over the prices that they will pay. If there are many buyers who are close rivals and yet differentiated one from another, each buyer will have individually only an insignificant effect on the market positions of rival buyers, but at the same time each buyer can individually exert some price control over the productive services that he buys.

Pricing of the services of material agents of production

The pricing of productive services purchased under conditions of monopsonistic competition is again analogous to the pricing of products in markets characterized by monopolistic competition. The differentiated buyer competing with many rivals has but a limited range within which he can control the price of the product he buys. Within this range he will maximize his profits by an input adjustment based on a comparison of marginal cost with marginal revenue; and marginal cost will be slightly higher than the price of a unit of input, so that this adjustment involves some curtailment of production. Since alternative buyers of these productive services provide very closely substitutive outlets, the supply of

the productive service to the firm will be very elastic; the discrepancy between marginal cost and the price of a unit of input will therefore be very small and the curtailment of inputs will be very slight.

The limits of the general range within which the single buyer can control price will be determined, again as in the case of monopolistic competition, by a rough relation between the aggregate demands of the entire group of slightly differentiated buyers and the aggregate offerings of productive services in the general market area. Increases in the aggregate offerings would lead to a lower general range within which each firm could manipulate the price of the productive services it buys. Similarly, increases in the aggregate demands of the group of differentiated rival buyers would push up the general range within which any one of them could manipulate price.

Some peculiarities of the pricing of services of labor

The generalizations just made concerning the pricing of productive services in markets characterized by monopsonistic competition are applicable for the most part to the determination of wages of labor as well, but there are some important differences between markets for labor services and those for the services of material agents of production. Aside from the greater importance of differentiation of outlets in the minds of workers, there is also the fact that the quality of performance is not entirely independent of the amount paid for the service and is more significantly influenced by the conditions of work. If an employer buys or rents a certain machine, the amount he pays the owner will have no effect on the performance of the machine. He will therefore try to get it for as little as possible. But under some circumstances the efficiency of each laborer will be affected by the wage he receives; and some employers will take this into account in their wage policies.

In view of the preference of workers for reasonably good conditions of work, we might expect that the employers who offered these superior working conditions would be able to take advantage of this fact and pay lower wages. In fact, however, the employers offering the best conditions of work are most frequently those also offering the highest wages. How may this apparent anomaly be explained? If it were just pure generosity most of these employers would not survive, although the greater efficiency of management in some cases gives a certain cushioning. The real explanation lies mainly in two lines: (1) These same employers frequently use higher wages as a basis for building morale and increasing the efficiency of workers employed; (2) they frequently hire workers not from the same homogeneous supply as those taken by other employers, but from a selected group of superior efficiency. An entrepreneur of superior efficiency will know how to use these more able men to get the most out of them, whereas another employer would be less able to realize the potential differences between the products of average and of superior workmen.

Long-run adjustments

Firms will enter or withdraw from a market in response to advantageous or disadvantageous situations in the buying of productive services quite in the same way as they will respond to conditions in the selling of the product. In fact, it is really taking into account the situations both in the markets in which a firm buys and those in which it sells that determines whether there will be profit or loss and hence whether new firms will enter or old firms will drop out. When firms move in in response to profit opportunities the result is an increase in rivalry in the buying of productive services as well as in the selling of products; the average expenditure schedule of a firm for a productive service will shift to the left, and the general level of prices paid for productive services by slightly differentiated buyers will rise, thus tending to check further profit opportunities. Conversely, when firms drop out, the remaining firms are more favorably situated than initially in the buying of productive services. These long-run adjustments among firms buying in a market characterized by monopsonistic competition are essentially similar to purely competitive adjustments except that the process is less exact and in a stable position some firms may still retain slight profits.

Appendix to Chapter 22

Graphic representation of long-run adjustments under monopolistic competition

The tendency to check profits where there are opportunities to produce substitute goods or substitute varieties may be clarified by a graphical

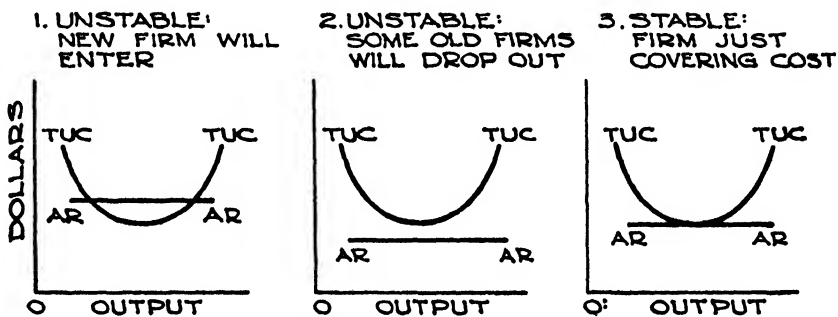


Fig. 22—A. Cost and average revenue of a new firm under pure competition.

comparison of these conditions with those prevailing under pure competition. Suppose that under conditions of pure competition selling price were temporarily higher than the lowest total unit cost of a new optimum scale firm; that is, average revenue, which coincides with market price, is greater than the lowest total unit cost of a new enterprise. If entry to the

industry is unrestricted, this will attract new firms, which will increase the output of the industry, lowering the price and hence the average revenue possibilities of each of the sellers in the industry. Conversely, in response to a situation in which average revenue is less than lowest total unit cost, some firms would in the long run drop out, and price and hence average revenue would rise. The short-run average revenue curve of a purely competitive seller is thus in a stable position only when it is tangent to total unit costs. In the long run, profits will tend to be eliminated. These re-adjustments of average revenue under conditions of pure competition with free entry are illustrated in Figures 22—A1, 22—A2, and 22—A3.

Similar adjustments will take place under conditions in which products are slightly differentiated and many new varieties are possible. If existing firms are charging prices such that a new seller of a differentiated product anticipates the costs

and the average revenue curve drawn in Figure 22—B1, he will anticipate a profit and there will be an incentive to set up in business. As more firms producing competing goods appear on the market, they draw customers away from existing firms, and the average revenue curves of each and all tend to shift to the left. During these adjustments the average revenues of some producers may have been pushed back until they are less

than enough to cover costs at any possible output; such producers must eventually either change their policies in some respects or drop out entirely. The long-run equilibrium position finally attained will resemble that under pure competition with freedom of entry in that (1) new entrepreneurs respond to profit opportunities in the general area until all such opportunities are being exploited, and (2) no firms remain operating at a loss. It differs in two important respects: (1) There is no tendency to adjust to a least-cost-combination output in each firm, and (2) profits persist in many firms except when there is "freedom of entry to the area."

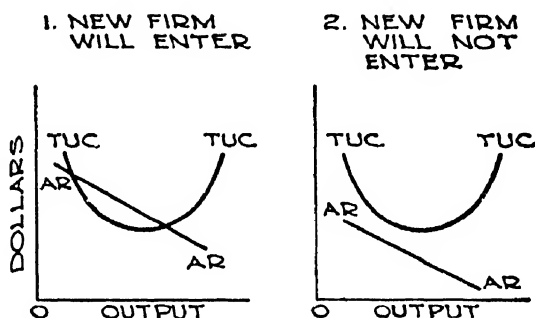


Fig. 22—B. Cost and average revenue in the most profitable opportunity for a new seller of a slightly differentiated product.

Long-run elimination of profits with freedom of entry to a "production area"

The concept of "freedom of entry" has no precise meaning except as it involves freedom to duplicate production processes, costs, and final products. Economists sometimes speak, however, of freedom of entry into production of a new variety of a group of products, even though there is not freedom to duplicate any single existing form. Freedom of

entry in this sense, which we may term "freedom of entry to a production area," would involve: (1) opportunity to produce a new variety that would place the new seller, once established, in the same revenue position as that in which his rivals would find themselves, and (2) opportunity to compete under the same cost conditions as were available to his rivals.

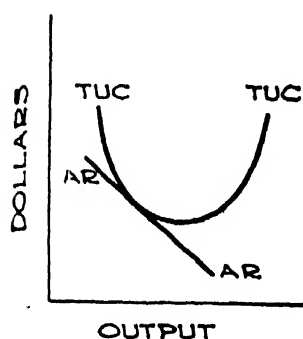


Fig. 22—C. Firm producing differentiated product just covers cost.

Where there is an approximation to such freedom of entry to a production area, profits will tend to be eliminated.

If there were such freedom new firms would continue to enter, pushing back the average revenue schedules of existing firms. Since under this assumption each enterprise is in the same cost-revenue position as each other enterprise, and profits among established firms indicate profit opportunities for new firms, new firms will continue to appear until each seller in the area is situated as in Figure 22—C.

For both old and new entrepreneurs average revenue is tangent to total unit costs, it is just possible to "break even," and profits are zero. Obviously, also, prices will be higher than lowest total unit costs, and each firm will be operating short of its least-cost-combination output.

CHAPTER 23

Oligopoly and Oligopsony

WHEN there are just a few rival sellers, each seller takes into account the ways in which rivals may respond to any price policy he may pursue; when there are just a few rival buyers, each buyer similarly takes into account the effects of his actions on the behavior of the other buyers. These situations are characterized as oligopoly and oligopsony respectively. The purpose of the present chapter is to examine briefly some types of behavior that may be expected in markets approximating these models of oligopoly and oligopsony.

In some concrete situations there may be several firms in an industry, but one of them so dominates the others that are very small and cluster around the leader that the adjustments in the dominant enterprise resemble far more the model of simple monopoly or monopsony than they approximate the oligopoly and oligopsony models to be discussed here. In the present chapter discussion will be limited to a consideration of cases in which there are several "important" firms, each of which may alone exert a significant influence in the total market for a product or a productive service.

Price Competition

"Nonaggressive" price competition among oligopolists

The purely competitive or the monopolistically competitive entrepreneur competes "aggressively" by cutting price in order to get more sales without paying any heed to responses of rivals, since such responses are either absent or negligible. What rivals do will depend on the general market situation and not on the actions of any single seller; they respond more or less simultaneously to these general market conditions, lowering prices when sales are falling off and vice versa.

The oligopolist entrepreneur on the other hand does not cut prices in such direct response to reduced sales; he will consider the effects of his individual price policy on the actions of rivals, and such considerations will give him pause. Rivals are likely to "meet" his cut, so he cannot anticipate attracting their customers away from them. Let us suppose firm A to be one of a small number of firms selling shoes. It may be that the situation is one of homogeneous oligopoly (the other firms are offering

an identical product); or it may be characterized by heterogeneous oligopoly (the offerings of any one firm being slightly differentiated from the others). In either case, firm A would not "rationally" ignore the reactions of rivals when considering lowering the prices at which it offers shoes for sale. Firm A would take into account the fact that firms B, and C, and so on, would probably "meet" this competition by lowering their prices, with the result that A's sales would probably be increased by only a very small amount relative to the increase that would follow if rivals did not thus respond. Firm A will therefore be far more reluctant to cut the price of its shoes in order to increase sales than if these other firms were expected to maintain their original prices independent of A's behavior. On the other hand, if in a market of only a few sellers one firm were to increase the price at which it would offer shoes for sale, there is a good chance that others might follow suit; the initial price increase by firm A would then result in a smaller loss of sales to A than would occur if the other firms failed to respond. There will be much more incentive to the individual entrepreneur to increase price if there is this possibility that his price increase will lead other sellers to raise their prices also.

Typically under conditions of oligopoly firms hesitate to reduce price except to "meet" the price cut occasionally introduced by one of the group; and such initiation of price cuts is therefore very much the exception. If he does initiate a price cut, the oligopolist usually proceeds not in the anticipation of getting trade away from rivals (unless possibly for a very short interval of lag), but rather in the anticipation of a general increase in the trade of all the firms in the group following a general "meeting" of his lower price. The focus of oligopolists on merely "meeting" the prices of rivals is a very negative sort of price competition; it is "nonaggressive" in contrast to the "aggressive" price policies of purely competitive and of monopolistically competitive sellers.

It is obvious from this discussion that the price that will prevail in an oligopoly situation will typically be higher than that that would prevail if these sellers were to engage in active price competition, ignoring the effects of their actions on rivals. This is true even though the different firms may not have got together in any formal planning of price policy. There is no clear-cut answer, however, to the question as to how high will be the price of a product in case of homogeneous oligopoly or the prices of the varieties of product in the case of heterogeneous oligopoly. This will depend on how rivals respond and how each firm evaluates the responses of others. If price is "too" high, the pressures to increase output will be so great that some seller will probably cut price and others will follow suit. If demand for the product or group of products is very elastic, output restrictions become severe with only a slight price differential above that that would be attained by aggressive price competition; if demand is very inelastic, rather high prices may be maintained much more easily. If new enterprises can make their appearance easily (which is unlikely) this possibility will influence short-run price policies that will be much more moderate than otherwise. If the members of an industry

have developed over a period of years a customary price, resistance to changes from this price will be great whether the equilibrium price that would result from aggressive price competition rises or falls. If the industry has experienced in the past violent oscillations of good and bad fortune, pressures to refrain from price-cutting may be much greater than otherwise. If firms have come to look upon one of their number as a leader in setting prices they may consistently follow the policy of this firm, usually a relatively moderate policy, established with a view to holding without break the position of leadership. Policy will also depend on pressures of public opinion. These and many other factors enter the situation. The only definite conclusion is this: *Where there is oligopoly, price competition will tend to be "nonaggressive," price (or prices) will therefore be higher than otherwise. Associated with this higher price (or prices) is restriction of production in each firm, and a smaller output of the industry or production area as a whole than would otherwise appear.*

Associated with the nonaggressive form of price competition in oligopoly situations is frequently a tacit policy for sharing the markets that offer restricted aggregate sales opportunities as a result of the maintenance of high prices. Conventions may be built up in an industry so that firms pursue a "live and let live" policy, refraining from intrusions into markets that have come to be regarded as the especial property of a particular enterprise, and refraining from efforts to obtain any appreciably larger share of a market than the firm has customarily commanded. This leaves room for some small degree of price-juggling where trade seems to be shifting significantly away from a particular firm; rivals may tolerate slightly lower prices without "meeting" them in such situations. But this tolerance of a slight price differential between firms is not inconsistent with the maintaining of a high general level of oligopoly price in the industry or "production area"; if the low-price firm were to cut below the margin of tolerance permitted in line with conventional market-sharing, rivals would respond by cracking down on their prices also.

Oligopoly collaboration in price and output policies

Thus far each firm has been viewed as acting individually. Frequently when there are only a few producers of a commodity, or of a range of varieties of a commodity, they get together, setting up agreements of a formal or an informal nature. In fact, there is scarcely any distinction between the tacit informal agreement of a group and the behavior of individual oligopolists taking into account the reactions of rivals. There are two major types of agreement, though the two are very frequently combined and are always closely related to each other: (1) The group may agree to follow a price leader or to maintain price at some given figure, accepting the individual output restrictions necessarily resulting from the high price. (2) The group may agree on some type of output restriction scheme as a means of maintaining price at a relatively high level. Frequently a price agreement is specifically implemented by an output- or market-sharing agreement.

The analysis of nonaggressive price competition among individual oligopolists provides the main clues to the functioning of price agreements, formal or informal. Such agreements generally set minimum prices, members of the agreement being quite free to raise prices if they wish to do so. Such agreements generally involve higher prices than would otherwise prevail, and with these higher prices are of course associated a smaller output of the industry as a whole and restriction of production by each member of the agreement. In practice many techniques have been developed to enforce compliance. The most serious short-run problems that must be faced by leaders or groups formulating collaborative price policies are conflicts with regard to the division of sales as among the different firms, and conflicting interests with regard to the most profitable price. Such conflicts have often led to a break-down of an agreement and are likely to be most serious when there is no associated specific provision for division of sales.

Groups of producers have developed a number of schemes for controlling outputs as a means of maintaining higher prices. The simplest type of control is to reduce the total quantity that may appear in a given market, dividing this quantity by some type of quota agreement among the members of the group. Under such a scheme, price may be left free to find the level at which the total quantity offered for sale by all firms together would find purchasers; but more frequently output agreements are a means of arriving at a workable division of the limited total sales resulting from a price agreement. They are therefore called "market-sharing" schemes. Although they cannot eliminate completely conflicts of interest between firms, such programs may serve to hold a group together for prolonged periods. Since the purpose of such agreements is usually to maintain prices higher than would otherwise prevail, higher price and smaller output will of course be the usual result.

Nonaggressive price competition among oligopsonists

Sellers in purely competitive and monopolistically competitive industries seek to attract customers by active price competition; oligopolist sellers usually refrain from such aggressive tactics. Similarly, buyers in purely competitive or monopsonistically competitive markets seek to attract more productive services by active bidding up of the prices of these services, that is by buyers' price competition; oligopsonist buyers usually refrain from such aggressive tactics. Each oligopsonist recognizes that his actions will cause rivals to change their policies, that if he bids up prices in order to get more hogs, or more services of labor, rival buyers of the hogs or the labor are likely to "meet" his prices. He will therefore be unable to attract these productive services away from rivals in any significant quantities; the increase in inputs that may be possible to him through an increase in the prices that he pays involves simply his share in the generally more available hogs or labor when all the oligopsonist firms are offering more for these things. The oligopsonist is thus "non-aggressive" in his price policies, refraining from any active bidding up

of prices aimed at getting a part of the markets for productive services away from rivals. This parallels exactly the nonaggressive behavior of oligopolists. Under conditions of oligopsony the prices of productive services will therefore be less, and the use of these services will be less, than would occur if the firms were to compete aggressively against each other. Oligopsony is also associated with conventional and nonaggressive sharing of the markets in making purchases.

Oligopsony collaboration in price and input policies

Collaboration among oligopsonists in buying is quite as important as the agreements among oligopolists in selling. Such agreements may again focus primarily on price, or on sharing of the market, or they may undertake a coordinated program of price and market-sharing together. When prices are held down through oligopsony agreements, the total available productive services will tend to be less, and the firms involved will be facing continuous pressures that may threaten their low price policies. Those who find themselves most hampered by the limited amounts of hogs or labor services available may break over and bid prices up in an attempt to improve their buying positions relative to other firms. In order to prevent such sporadic actions, which may eventually undermine the oligopsony price control, the group may arrive at an agreement on the basis of which the limited quantity is divided up among the firms. They may share the market according to purchases over some preceding period of time or according to some other criterion on which the group manages to agree. Such market-sharing agreements among buyers obviously facilitate the maintenance of agreements designed to keep prices down.

Crudeness of approximations to maximum profit marginal adjustments in a firm

When there are only a few rival firms in a market, whether they are rivals in the sale of a product or in the buying of a productive service, many complex considerations will influence the policies of each of the entrepreneurs. All would like to maximize profits, but how can this be done? What kinds of anticipations of rival reactions will be taken into account? It may be that each firm can vary its competitive policies within limited ranges without evoking any reaction, behaving essentially as a firm in a market characterized by monopolistic or monopsonistic competition within limited ranges in price and quality competition; yet the same firm may function as one of an oligopoly or oligopsony group so far as more drastic shifts in policy are concerned. It may be that some kinds of policies, as price-cutting, will generally bring a response from rivals, and will therefore be undertaken with great reluctance and only after extended deliberation; other policies, perhaps quality or advertising competition of one form or another may be undertaken by the same firm without evoking a reaction. Moreover, there is a political power game that may go on within the group of dominant firms, a shuffling for position and relative privilege within a general program of group action.

This group action may involve tacit agreements to refrain from active price competition, and to share the markets involved according to customary relative positions of the firms; it may be expressed in more formal arrangements.

Yet in all these variations of policy and political bargaining positions as between rivals, there is a certain crude approximation in each firm to the comparison of marginal revenue with marginal cost that was discussed in Chapter 12 as characterizing all monopolistic and monopsonistic markets. Each entrepreneur, in the light of his interpretation of the reactions of his rivals, will consider whether all things taken into account he would add more to his revenues than to his costs by increasing his input-output level, whether he would reduce costs more than revenues by restricting production. Since he plays a significant role in the total market the oligopolist is in a position to influence the price of what he sells, and similarly the oligopsonist is in a position to influence the price of what he buys. The oligopolist can only increase significantly the sales of a particular variety of a product by taking a lower price, the addition to his revenue (marginal revenue) will therefore be consistently less than the sale value of the marginal product, and he will be inclined to curtail production. The oligopsonist can only increase inputs significantly by paying a higher price, the addition to his cost (marginal cost) will therefore be consistently greater than the average expenditure per unit of input, and he will be inclined to curtail inputs. In view of the complexity of rival reactions, entrepreneurial estimates of the cost and revenue possibilities are inevitably less precise than in other types of situations. Nevertheless, it is at the maximizing of profits through adjustments in line with marginal costs and marginal revenues that such entrepreneurs are aiming, and toward which their actions will be directed. Oligopolists will tend to curtail production in order to hold up the prices of the products they sell; oligopsonists will curtail inputs in order to push down the prices they pay for productive services. Adjustments in either case or in a combination of the two will involve curtailment of production short of the level that would be attained if comparisons were made between sale value of marginal product and average expenditure per unit of input.

The shunting of competition into non-price channels or indirect price competition

Wherever rival sellers are especially sensitive in their responses to the policy of a firm, as is most strikingly the case in price-cutting, the entrepreneur seeking to gain more customers will find other methods of attracting them that may not evoke so immediate a reaction from rivals. Where the absence of active price competition is further emphasized through formal agreements this tendency is of course further exaggerated. The result of oligopoly restraints on pricing practices is therefore to encourage more active use of other techniques of competition through quality differentiation and advertising, or through indirect forms of price-shading which

may "get by" without evoking immediate retaliation. In some general areas of production the differentiation of sellers may provide a sufficiently satisfactory solution to the problem of sharing the market so that oligopoly prices can be maintained quite successfully in the absence of any other form of market-sharing or output-restriction agreement. This emphasis on indirect methods of price competition, on quality competition (including both form of product and conditions of sale), and on advertising competition, has tremendously important implications for planes of living and general economic welfare. These forms of competition will be examined in the next chapter. For the present they will be ignored.

Long-Run Adjustments

Long-run adjustments in oligopoly and in oligopsony situations may result in the sharp curbing or elimination of profits; on the other hand profits may be maintained for extended periods of time, depending on whether entry into the industry or production area is essentially free or is sharply restricted.

Long-run adjustments with freedom of entry

If entry into an industry characterized by oligopoly is free, new firms will appear so long as profits can be obtained. As these new firms appear they will share the markets with the previously existing enterprises, and each of the initial enterprises will suffer a reduction of sales volume. The new firms, once established, will usually join in the nonaggressive price policies or the price agreements of the initial members of the group. Under continued pressures reducing sales per firm, prices may be lowered slightly; but it will not be primarily the lowering of price that will eliminate or curtail profits. The oligopoly situation is in sharp contrast with pure competition and in partial contrast with monopolistic competition in that the primary factor reducing profits is the reduced sales per enterprise instead of the cutting of price.

Because price in oligopoly situations is maintained at high levels through agreements or nonaggressive price policies, firms in the industry or production area will continue to make profits in situations in which aggressive price competition would otherwise eliminate these gains. As a result, there will continue to be an attraction of new firms into the industry after existing firms are already adequate to fill demands. This continued entry of new firms will lead to additional investments in relatively fixed agents of production, and each firm will operate at smaller and smaller production levels. When an equilibrium position is finally attained price will still be high and therefore aggregate sales will be short of what they would be if price competition were more active; there will be more firms in the industry or production area than would be needed to perform efficiently the services of meeting consumer demands; sales of each firm will be so small that profits will be negligible in most cases and there will be no profit opportunity for further new enterprises.

Long-run adjustments in oligopsony situations parallel those in oligopoly. Nonaggressive price competition among oligopsonists leads to the maintaining of low prices; and firms buying in these markets will therefore continue to make profits in situations in which aggressive price competition in buying would otherwise eliminate these gains. As a result there is a continued encouragement to new firms to enter into the oligopsony market even after there are already a sufficiency of firms to carry on efficiently the same or a greater aggregate level of productive activity. Adjustments to the entry of firms tend to occur primarily through the dividing of the market into smaller and smaller parcels per firm instead of the bidding up of the prices of the productive services. The result is again the eliminating or curtailing of profits, with small inputs per firm, small aggregate purchases, and low prices for productive services.

Restrictions on entry

Restrictions on entry of new firms into markets controlled by a small group of firms are very common. These restrictions take much the same form as in the case of monopoly and monopsony. They may center in special privilege positions of the "insiders," in group control over crucial patents, in group use of tactics of unfair competition against outsiders, and so on. Oligopoly situations in selling are frequently supported by exclusive controls of the group in the oligopsonistic buying of essential productive services, and vice versa. Wherever restrictions on entry can be effectively maintained, the small group of firms may obtain profits over long periods of time. Even where the bases of exclusion can be broken down they may serve to protect profit positions for relatively prolonged periods.

Appendix to Chapter 23

Graphic presentation of aggressive versus nonaggressive price competition

A graphic illustration may serve to clarify further the distinction between aggressive and nonaggressive price competition. When there are only a few sellers it is unrealistic to draw for any one of them an average revenue curve based on the assumption that other things remain unchanged. Some understanding may, however, be gained by comparing the way in which a seller would behave when he takes into account reactions of rivals with what he would do if, though these rivals will in fact react, he assumed that they would not respond to his policies. In fact, in many concrete situations sellers are "irrational" to the extent of disregarding pertinent considerations of rivals' reactions. For such a seller two sets of estimated average revenue curves may be drawn. One of these indicates the amounts which the seller estimates he will be able to sell at a series of prices after taking into account his expectations as to how other sellers readjust their policies in response to his. *On the assumption that his*

estimates are correct, these revenue curves may be termed the seller's "real" demand (*rd*) and "real" marginal unit revenue (*rm*) curves. The other set of revenue curves is completely "imaginary." These are the revenue curves estimated on the unreal assumption that other sellers would not change their present policies in response to the changed price policies of the individual seller, that is, on the assumption that all "other things" remain unchanged. These estimates are "imaginary," since a shift from one price to another will in fact change the "other things." If nevertheless a seller were to act as if this "imaginary" curve were "real," price competition would be much more effective, price would be lower and output greater than when he takes rivals' reactions into account.

In Figure 23—A, there are two average revenue curves for a firm X.

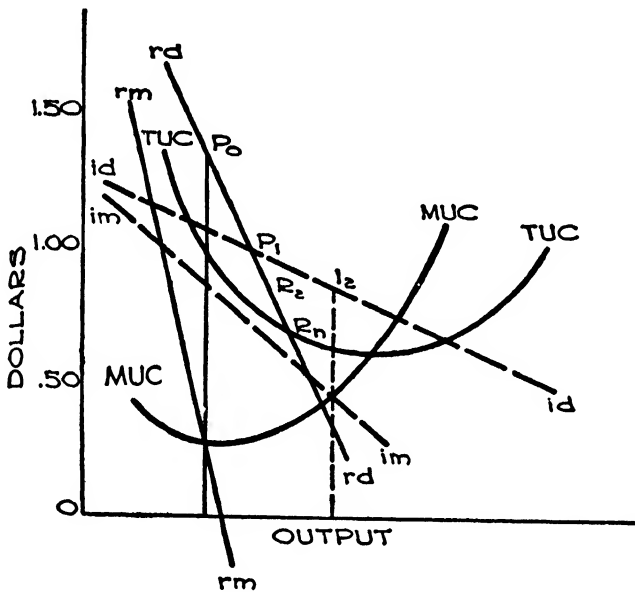


Fig. 23—A. Nonaggressive price policy.

The curve *id-id* is the imaginary average revenue curve and *im* the marginal unit revenue curve derived from it. The curve *rd-rd* is the amounts that this producer would estimate that he could sell if he takes into account the actual price and other re-adjustments of rivals (*rm* is the marginal unit revenue curve derived from *rd*). The actual initial price of \$1.00 (P_1) is the same on the two curves, but all other points differ. A seller ignoring rival reactions in this situation might try cutting price to, say, \$.90, anticipating an increase of sales to I_2 ; but rivals would cut their prices in response and his sales would increase only to R_2 . Had the "imaginary" schedule been "real" this change would have been profitable, but in fact it was not. Conversely, this seller would not raise price to \$1.10 if his decisions are based on the imaginary curve, since in the absence of rival responses his sales would practically disappear; but in

fact others may raise their prices too so that the decrease in his sales would be very slight and profits would be greater than before.

This little demonstration shows clearly that in those situations in which sellers ignore the reactions of rivals a much more aggressive price-cutting policy may be expected. This is the "rational policy" when rivals in fact will not react to policies initiated by a single firm. But in the case in which the demand curve assuming no rival responses is imaginary, a new situation arises. *Each shift in price downward will require the construction of a new imaginary demand curve passing through the new real point.* If these imaginary curves are all sufficiently elastic and the seller acts on the basis of the estimated imaginary curve, price may be pushed by aggressive price competition down to or even below R_n , where costs are just covered. It is evident, however, that in practice no such seller would be likely to ignore the reactions of rivals to this extent.

CHAPTER 24

Non-price Competition and Price-Control Evasions Among Sellers

DIRECT price competition is but one of many ways, and not necessarily the most important way, in which rival sellers seek to gain customers from each other. Offerings of stronger and more durable products, of credit and delivery services, of guarantees, of attractive surroundings in the selling of goods—appeals to taste, prestige, fashion—full-page advertisements with pictures of pretty girls, better radio programs, neon lights and trick electric signs—these and many other techniques are employed in the competitive game. Arguments as to whether competitive behavior in general is increasing or diminishing are inconclusive, but there is no doubt that through the 'twenties and 'thirties trends were evident to indicate a shift away from price competition toward an increased emphasis on the non-price aspects of competition.

Commenting on the social importance of non-price competition, the authors of the TNEC monograph "Price Behavior and Business Policy" write as follows: ¹

Non-price competition is of particular importance to the standard of living of consumers because of the extent to which it affects retail markets and the everyday necessities of life. The amount which a family must spend for food, clothing, groceries, drugs, and cosmetics is related to the manner in which business concerns selling these products choose to compete, by their decisions to stress or skimp quality, to advertise more or less intensively, to pack simply or elaborately, to favor or oppose retail price-cutting, and so on. This is clearly a matter of broad public concern.

In previous chapters brief attention has been given to some of the forms that non-price competition may take. These were roughly divided into two groups, quality competition and demand creation.

Quality competition was used very broadly to include a wide range of competitive practices that may be roughly divided into two subgroups: (1) the physical quality of a product in use, that is its durability, adaptability, aesthetic appeal, and so on; and (2) the combination of associated services that accompany the sale of the product, including packaging,

¹ TNEC Monograph No. 1, p. 59.

terms of payment, delivery, atmosphere of the shop, and so on. Only the first of these subgroups is generally regarded as "quality competition proper." The second group merges over imperceptibly into advertising competition or demand creation. It is theoretically distinguishable from advertising competition according to whether it is designed as response to existing demand or as an attempt to change demand, but in practice this distinction is sometimes difficult to make.

Demand creation is also a broad and inexact concept. It includes the spotlighting of brand names and trademarks as means of attaching consumers to a particular product. It includes the many techniques of persuasion. It seeks both to inform and to misinform in order to attract customers. It is a tool for bringing to the attention of potential consumers the quality or price offerings of the seller.

Where direct price competition is curtailed, competitive activities are redirected into other channels such as quality and advertising competition. There are also efforts to find rather direct substitutes for open price competition through the use of premiums, allowances for trade-ins, and a multitude of other techniques for indirectly shading prices. The significance of these devices is often underestimated.

Quality Competition

In a very simple economy, such as that of the early colonial days, the variety of goods was small and buyers were relatively well informed concerning ways of judging quality. In these circumstances the elaborate conditions now surrounding sale of products, especially at retail, were unknown. Informed buyers were concerned over quality and quality competition was probably very much more effective than it is today. The spirit of craftsmanship lent added effectiveness to this competition.

In modern times the situation is very different. New techniques of production make deception in quality much easier and much more difficult to detect. Higher planes of living have involved also a diversification of products almost inconceivable. The results are on the one hand much greater complexity in the problem of selection facing buyers, and on the other an opportunity to use fine distinctions of product in place of price-cutting as a means of attracting customers. Price comparisons are under these circumstances very poor guides in purchasing.

The problem of measuring quality

What is high quality? This is a very complex and difficult question; it is a question that consumers try to answer in relation to their particular preferences; it is a question that both private and public agencies must wrestle with in attempting to provide information to consumers; it is a question for statisticians and economists seeking to study and measure the changes in quality over time and the functioning of price and quality competition in any given product. Even confining attention to what has been designated as "quality competition proper," to the physical

quality of goods in their effectiveness in use, the measurement of quality is extremely difficult.

What is good quality in a radio? Is tone, or selectiveness as between different broadcasting stations, or distance, or durability, or low cost of maintenance, or aesthetic design, more important? How may each of these things be measured and how may their relative importance be determined? Deciding on quality in automobiles is yet more difficult. Even for relatively simple products such as coats, suits, dresses, the problem of judging quality is a baffling one. Some characteristics of a good may be expressed in quantitative terms, as gasoline consumption of a car, estimates of cost of upkeep of a radio, tensile strength of steel bars and of cotton sheets. Others are much more elusive, as style in clothes, aesthetics and style of chromium trim on automobiles, and so on.

Range and emphasis in quality competition

The range of quality competition is clearly great. It extends all the way from the relatively measurable characteristics of durability and operating economy through the immeasurables of design and exclusiveness and the associated services, packaging, and so on that are provided with the sale. The extent to which any of these factors are emphasized in competitive strategy will depend on conditions in the particular industry. Where purchasers are expert and interest is primarily in utilitarian performance emphasis will be on the relatively measurable aspects of quality; this is most likely to occur where buyers are industrial users of the product in further production processes, and in some cases of government buying. Where purchasers are most interested in factors such as style, quality competition will be channeled in these directions and much less attention will be paid to durability and strength. And where purchasers are inexpert, as many of the ultimate consumers of goods sold at retail, sellers take advantage of this situation to compete through the relatively superficial aspects of quality that catch the eye and through emphasis on associated services and convenient conditions of sale.

Those aspects of quality improvement that are readily measurable are very like price-cutting as competitive devices for attracting customers. While they may be effective in increasing sales, they may also be duplicated by other sellers. The fear of retaliation from rivals may discourage aggressive competition in these aspects of quality. They are also the most readily standardized and the most likely to be made the object of standardization agreements among competing sellers. As a result there is an increasing tendency to emphasize the nonmeasurable aspects of product differences. Variations of this sort can be developed ad infinitum; they will not be exactly duplicated by rivals, and they may be quite effective means of holding customers.

Closely associated with quality competition is the use of guarantees. Such guarantees with respect to the relatively measurable aspects of quality, as the performance of automobiles, washing machines, radios, are also means for diversifying appeals. Guarantees to industrial buyers

that prices will not be cut over some stated future period, or that if they are the buyer will be rebated, have the effect of reducing price flexibility. Most guarantees give a competitive advantage to the concerns first initiating them; but most of them can also be easily matched. Once they have become general in an industry, therefore, new methods of competing will be sought.

Demand Creation

The analysis of prices and quantities of goods thus far undertaken has proceeded on the assumption that sellers exercised no direct influence on the demands of consumers for their goods. The fact that consumers may frequently be misinformed in their selections has been largely ignored. Advertising competition is a technique by which a particular entrepreneur seeks to attract customers. He may succeed in part merely by informing them of opportunities of which they were not formerly aware; he may succeed by misinforming them concerning what he has to offer as compared with the offerings of other sellers; and he may persuade them to change their underlying preferences in favor of the kind of thing he is offering for sale. Any realistic understanding of the functioning of economic processes requires an examination into the nature and effects of this powerful competitive technique.

Increasing the demand for the firm's product by advertising

Since the intent of a seller in advertising is to attract more customers to the purchase of his product, he will try to impress upon them the superiority of his product in relation to the price at which he is offering it. If he succeeds he will be able to sell more units at any given price, or to get a higher price for any given quantity. The demand for this particular product (the seller's average revenue schedule) will have increased. He may attract customers by setting his product apart from others in their minds, as distinctly superior and therefore worth a higher price. Insofar as he thus increases the extent to which consumers differentiate him from other sellers (whether this differentiation is real or illusory) he obtains for himself a greater degree of control over price. He can raise his price above that charged by rivals and yet retain a goodly number of customers. On the other hand, a seller previously charging exceptionally low prices may lead consumers to believe that his product is just as good as that sold at a higher price by rivals; and whether or not this is true he will be able in this way to get more customers at existing prices or to increase his price and still retain the original number of buyers, whichever he may see fit.

The meaning of "selling costs"

Costs of production may be defined as those costs incurred in creating and adapting a product in response to given demand conditions; selling costs are costs incurred in attempting to shift the demand to the advantage of the seller. Thus manufacture, transportation, storage, provision

of credit, delivery, convenience of location—all these are production costs. Newspaper advertising, window displays, sales departments insofar as they deal in persuasion but not in physical handling of goods—these are selling costs.

The distinction between production and selling costs is important for the fuller understanding of price-making. When all costs are production costs demand is determined independently of costs. By analyzing the relation between cost and revenue data it is possible to arrive at a solution to questions concerning price and output in individual firms and in an industry. When one part of costs is directly related to demand a new element appears in the problem. There is the possibility of various total outputs not only at different prices but also at the same price with different selling costs.

If it were possible to determine exactly what would be the effect on demand of a given expenditure on selling efforts or of a given increase in such expenditure, the entrepreneur of a firm could estimate the maximum profit adjustment taking these costs into account. Each concern is in a position to change price and output with any given demand, to change quality, and to change demand. For each quality of good and expenditure on selling efforts there will be an associated average revenue curve and a maximum profit price-output adjustment. The producer will experiment and estimate the various possibilities, aiming at a situation in which his profits are the highest possible, in which no change available to him could increase them.

Sellers' difficulties in estimating the effects of advertising

Estimating future costs and future demand even in the absence of advertising is sufficiently difficult, although these things can usually be approximated to a fair degree of accuracy. But selling costs introduce still further difficulties.

There are many reasons why it is by no means easy to determine the effects on demand of any given advertising program. There is a less consistent relation between selling costs and sales than between production costs and output. Large enterprises spend huge sums annually not only directly in advertising, but in surveys to determine the effects of particular advertising programs. No conclusions can be drawn merely because sales increased when a particular advertising expenditure was made, since other developments may be the true explanation of the sales increase. Moreover, an advertising technique that is new may be very effective for a time, but lose its effectiveness when people get used to it or it is adopted by other sellers. Some advertising becomes necessary to "meet" that initiated by others in order to maintain anything approaching a satisfactory position in the market.

Advertising differs from price and quality policies in that action at any given time is more likely to have a relatively enduring effect without repetition of that action. Every firm attempts to obtain the "good-will" of a group of customers who will continue buying in the same way even if

the advertising is dropped. One of the most difficult questions for the advertiser is therefore to know when he is in this position—when he can let up on his advertising outlays (selling costs) and still maintain his position in the market. How is he to know how many of his customers are “repeat” customers, and how is he to know how easily repeat customers can be induced to trade elsewhere if he relaxes his advertising efforts? He knows only that if he once allows repeat customers to be diverted to other sellers he will find it far more costly to win them back than he would have found it to retain them in the first place. These uncertainties result in some reluctance on the part of entrepreneurs to experiment with reducing selling costs even though they may suspect that to do so would increase profits.

Competitive demand creation

Competitive demand creation is most evident among rival producers of only slightly differentiated consumers’ goods, for example different brands of cigarettes; but there is also advertising competition between producers of candy and flowers, and more broadly between producers of automobiles and “all other things.”

When a seller engages in a new advertising expenditure he expects his market position to become more favorable to him, he expects that at any given price and quality he will be able to sell more than he otherwise could. His gain takes place largely by drawing customers away from other sellers who are in various degrees of rivalry with him. How long his improved market position will persist will depend on the one hand on consumer preferences and the skill of the advertiser in getting something that “clicks.” It will depend on the other hand on whether his rivals in turn draw customers away from him. This they might have done anyway through independent initiation of advertising; but his action may directly stimulate them to engage in active advertising competition. They may do this simply in imitation of him, or they may try directly to “win back” the customers he has drawn away from them.

Let us suppose first that an advertiser of a particular brand draws just a few customers away from each of many rivals. It may be that no one of them will consider retaliatory advertising to win back these few customers; but even so, many of these rival producers, observing the success of the initiator of the advertising, will decide to imitate him. Since they are not first in the field, their efforts cannot have exactly the same effect. The process may be slow, and *during the lag the initiator of the advertising competition will gain a more profitable demand position*; but eventually his exceptional gains are likely to be eliminated by the manifold new advertisers taking his trade. Their imitation of his advertising may have neutralized the effects he at first realized, and the demand for his product may be back where it started. He cannot return to his previous cost situation, however, since if he were now to cease his advertising expenditures rivals would draw still more customers away from him. He must keep on advertising in order to maintain his posi-

tion in the market, and he may in fact be under pressure to exert some new advertising effort in order once more to get at least for a time an especially favorable position.

If instead of drawing customers from many rivals an advertiser draws them heavily from a few, the response of these rivals would be more certain and more immediate and direct. They would retaliate by trying to get their customers back again. Fear of such retaliation may act as a check on the initiation of aggressive advertising, but it is more likely that it will be a relatively insignificant force. Each seller fears that his rivals may "get the jump on him" more than he fears the direct retaliation that will follow his initiation of advertising competition. There is always the possibility that he may hit upon something that will have quite a high degree of invulnerability to attack from rivals attempting to retaliate.

Some of the most successful advertisers have doubtless subscribed to the idea that "a good offense is the best defense." They would probably add that a dollar spent in an advertising offensive now will save nine dollars spent later in attempting to win back customers lost to some other more aggressive seller. After customers have once got the habit of patronizing a certain seller, it is very much more difficult to draw them to a new enterprise than before they formed any such allegiances. New firms may hesitate to risk such an undertaking, and the well-established firm may operate at high profits for a long time before rivals make significant intrusions into its market. The man who is "out ahead" in advertising skillfully is the man who will obtain these advantages, and it is this fact more than any other that leads sellers to seek continuously new and more effective ways of gaining a leadership position through the use of advertising techniques.

The incentives to advertising competition, both "offensive" and "defensive," are so great that in some areas of production there has been a continuously upward spiral of selling costs. A leader is imitated or retaliated against; his temporarily more favorable demand position disappears, but he must continue advertising to retain a place in the market; each seller again sees an opportunity to gain at least temporarily from yet further expansion of advertising, others "meet" his competition, and now on the new level with higher selling costs they again neutralize each other's efforts to win customers; again no one dares to withdraw his advertising, and each may temporarily gain by further extending it. So the cumulative spiral continues upward. There is usually an advantage in leading, great danger in curtailing. Only when they have reached extremely high levels have there been signs of concerted efforts among groups of producers to modify expensive advertising programs.

In view of this upward spiral of advertising costs, the effects of which are continuously in the process of being neutralized so far as each seller's market opportunities are concerned, it may seem strange that associations of entrepreneurs sometimes chide members of the group who have spent only small amounts on advertising. The explanation lies in the effects

of advertising in diverting consumers' dollars into buying more of a whole class of commodities, say more cars or more radios. Sometimes advertising is definitely set up in terms of a whole class of goods; but even brand advertising may have such general effects. The magazine advertisement for Fords may increase the interest of consumers in automobiles in general, and many of these same consumers may decide in the end to buy Plymouths. Only in exceptional cases (such as flowers versus candy) is the impact of such advertising likely to be sufficiently focused against any particular firm or any coordinated group of firms to stimulate directly either imitation or retaliation.

Advertising versus price-cutting and quality competition in business policy

Customers may be attracted to a firm by lowering price, by raising quality, or by increasing advertising expenditures. There are many situations in which increased advertising will be the method chosen. Positive pressures to do so are of various kinds.

1. For some products aggressive advertising is so extensive that a seller who does not compete in this way will find his sales dwindling to almost nothing simply because consumers are unaware of his existence. He sacrifices economies of larger-scale operations, a sacrifice that may be very serious. Without enough advertising to attract attention to his offerings, he would find price-cutting or quality improvement to be of no avail. Once he has met the advertising competition of rivals, however, further expansion may be accomplished by price and quality adjustments.

2. An entrepreneur is continuously making estimates as to future conditions and the effects of various alternative policies on sales and profits. If he lowers prices in order to attract customers he runs the risk not only that the experiment may fail in its purpose but that it may lead consumers to believe his product is inferior. Therefore, when he tries to raise his price once more, his sales may be less than before. If he had sought to increase sales by more advertising instead of by cutting prices, and the experiment had failed, he would have been able to go back to his original position without any loss of customers resulting from his actions. When advertising and price-cutting seem to offer equal chances of gaining customers at comparable costs he will, therefore, probably choose advertising, as involving less risk.

3. Another reason for choosing advertising rather than price-cutting or quality improvement as a means of increasing sales may be that with the use of advertising there is a greater chance for more permanent results from a temporary program. And even where results are only temporary, the lag in effective rival reactions is likely to be greater than in the case of price competition, and the interim period may promise attractive profits.

4. Finally, in oligopoly situations producers exert strong pressure on each other against price-cutting, pay less attention to quality competition, and sometimes definitely urge members of the group to advertise in order

to attract consumers from other general areas of expenditure into this one.

Escape Devices in Lieu of Open Price Competition

While to many businessmen non-price competition offers profit opportunities much greater than are available through open price competition, this is primarily true of large enterprises with established reputations and resources sufficient to put over big advertising campaigns. There are still many firms, especially new and small enterprises, that find their greatest opportunities in price competition. Frequently it is only through cutting their prices below those charged by the big companies that these small or new concerns can get any place at all in the markets. This situation is recognized in some industries by large concerns that permit small firms to undersell them by a slight margin without threat of retaliatory action.² Indeed in some industries it is more profitable to the big companies to refrain from a too complete and strenuous process of weeding out small firms than to engage in a long series of small local price wars.

The conflicts of interests between firms focusing on non-price competition and those making use primarily of price competition as means of gaining and retaining customers become very sharp when in periods of limited opportunity there is an attempt to coerce all concerns into conforming without exception to a general agreement as to prices. Under pressure in such situations, substitutes for open price competition will be sought, especially by those firms placed at the greatest disadvantage under uniform prices.

Conflicts of interest as revealed under the National Recovery Act

During the depression of the early 'thirties, sales opportunities were becoming generally less and less and under these pressures long-standing price maintenance agreements and tacit following of price leaders in oligopoly industries began to crack. As an emergency measure, an attempt was made to improve business conditions generally by a cooperative program between business and government that set up codes for the behavior of the members of each industry. Many of these codes included provisions setting minimum prices below which no firm was to sell. These provisions eliminating price competition worked great hardship on some firms, especially on small enterprises in markets dominated by a few large and well-known concerns, or where some firms offer services along with their sales while others do not. These difficulties are illustrated in reports and hearings:

In the Cleaning and Dyeing Trade the cash-and-carry cleaners contended that they could not keep their business unless they were permitted to sell below the prices of those cleaners who receive and deliver the garment at the customer's residence. Their refusal to abide by the uniform minimum prices established under the code was an important factor in the breakdown of the price structure. In the Wholesale Confectionary Trade a cash-and-carry wholesaler complained

² See TNEC Monograph No. 1, p. 51.

to a member of his local code authority that he must either allow a larger discount than this competitor who delivered merchandise and extended credit or else go out of business. As he later described the conversation: "The answer was this, 'Mr. Lieberman, you have been a cash-and-carry man. From today on you will become a service man.' I said 'You are a code authority lawyer today; what would happen if someone wanted you to become a criminal lawyer? Would you like that?'"³

The resentment of large manufacturers at price-cutting by small firms in markets in which they have dominant positions is illustrated at many points in hearings before the National Recovery Administration. They claim that they have built up the markets for their products, that they have created demand through extensive advertising for which they have incurred heavy expenses, and that it is unjust that newcomers should be permitted to "horn in" on their markets through price-cutting without incurring their share of advertising costs. On the other hand, it is just this action on the part of new firms that protects consumers against monopoly control. This situation was well illustrated in the tire manufacturing industry. The large manufacturers sought to establish and enforce minimum prices through the NRA codes. Here was the answer of a small producer:

We don't want to sell below cost. We don't propose to sell below cost, but you notice what they add to it. Nobody must sell below cost unless to meet competition.

Well, up here is a big manufacturer. Everybody knows he has to sell above us if he can make money. He can do it and has consumer acceptance that carries him on. He would starve if he sold at our price. There is no question about it. He does not have to. He keeps on advertising. Here is our cost. He is not required to sell at his cost when he "meets" our competition, so he can come down and meet us and we have not any right under the formula to fall below; in other words, as I have heretofore said, we are chained to a post, and he comes and gets us and sells against us as long as he wants to because of that provision. Don't you see where it all leads?⁴

These situations did not occur uniquely under the NRA but have appeared and reappeared wherever rigid controls have prevented open price competition, whether governmentally enforced or supported by pooling agreements among private concerns without government support.

Devices to evade price control schemes as revealed in NRA regulations

The use of indirect methods of evading minimum price regulations was very common under the NRA. Commenting on the evasions of minimum price agreements in the codes, the report of the Division of Industrial Economics makes the following remarks:

Wherever minimum-price systems cause peculiar difficulties for one part of an industry, there was a direct incentive to evade or destroy the minimum prices. Minimum-price provisions were consequently difficult to enforce.

³ Quoted in Temporary National Economic Committee Monograph No. 1, pp. 91-92.

⁴ *Ibid.*, p. 56

In the sale of tires prices could be cut by extending the duration of guarantees of performance. Price concessions in the retail sale of automobiles consisted largely of higher trade-in allowances upon used cars. In the lumber and bituminous coal industries, price cutters delivered better grades than had been ordered and paid for. In the bituminous coal industry some producers offered guarantees which their customers knew were excessive and contracted to forfeit a part of the price if the coal was below the guaranteed standard. In many industries premiums or free deals were used to reduce the net price. Some producers attracted trade by offering lavish entertainment to buyers. Others offered unusually liberal cash discounts and credit terms. Long-term contracts whose date was alleged to be prior to the minimum-price system served as excuses for sales at low prices to favored customers.

The range and variety of possibilities of evading price control is indicated by the long list of types of regulation undertaken by the NRA. This list includes the following thirteen main groups of concessions or methods of concealing concessions, and the variety of techniques listed under these thirteen heads includes 222 different items. All of these are familiar in business practice and the list might have been extended yet further.

Types of Concessions to Influence Sales as Dealt with by Codes

1. Concessions primarily related to time of buyer's payment
2. Concessions primarily related to risk of buyer
3. Concessions primarily related to supplying additional goods
4. Concessions rendered buyer through use of sellers' employees or property
5. Concessions granted buyer through financial assistance or favors
6. Concessions related to manner, and/or time of shipment
7. Concessions through payment or diversion of commissions or fees to customers
8. Concessions through acceptance of competitor's materials from buyers
9. Concessions through allowances or payments for value rendered by buyer
10. Concessions through sale of substandard or obsolete goods
11. Concessions granted during performance contrary to provisions of agreement
12. Acceptance of forms of payment in which concessions may be concealed
13. Types of agreements, offers, invoicing, and so on, by means of which concessions may be concealed

The comments in the report of the Division of Industrial Economics, and the implications of the long list of regulated practices are important not only as history of the experience under NRA but also as indicators of how businessmen will behave when price regulation is attempted. Such evasions were not new. They had appeared in many cases in which prices seemed to be fixed and held stable through oligopoly policy.

Analogous evasions of price ceilings

Analogous to the development of techniques of indirect price competition among sellers are devices to evade price controls in buying. Oligopsony groups seek to hold down the prices of the productive services that they purchase. Wherever there are very great profit opportunities, however, the incentive to bid up prices in order to get more productive

services becomes increasingly strong. In boom periods, and most strikingly in a war economy, the result is likely to be the cracking of oligopsony price "ceilings," and price jumps above these ceilings. Wherever such an agreement is strongly enforced, and especially where price ceilings are backed or even initiated by government, the increases in prices above the ceiling may appear in somewhat concealed ways. Perhaps buyers unable to get as much lumber as they wish will reach an agreement with lumber producers to deliver to them grade B lumber at grade A prices; this is satisfactory to both buyers and sellers, since the seller gets a bigger income and the buyer gets lumber he very much wants, which would not otherwise have been available to him. A popular story now familiar to many Americans is that of the evasion of price ceilings on hogs in Germany; it is said that the farmer sells his dog along with his hog, and no one is very much surprised when the dog runs back home. Whether this particular story is true or not does not matter; it illustrates the variety of human ingenuity in evading price controls where the incentive to do so is strong.

CHAPTER 25

A Philosophic Intrusion: Individualism and the Economy

THE philosophy of individualism that has been basic in American traditions focuses on the freedom of the individual in what he does and says and thinks. In economic life this philosophy has found expression in an aggressive private enterprise system. That system has been justified by many of its adherents as a social order which gave on the one hand *freedom to individuals as consumers* to choose what they wished and thus to command resources to be used in producing according to their preferences, on the other hand *freedom to individuals as producers* to engage in the businesses or occupations of their choosing. To what extent the private enterprise economy in the United States has in fact realized these results is frequently debated; but though no complete quantitative agreement is likely to be reached, students of that economy agree concerning the *kinds* of maladjustments that make their appearance in that setting. A preliminary examination of various practices now common in the American economy in their effects on resource allocation and on freedom of opportunity will be undertaken in this chapter. In some respects it is therefore just an extension of the discussion of Chapter 20; but it goes far beyond that chapter in scope and in the variety of problems considered. Finally some conclusions concerning the evolution of American individualism and its implications and limitations will be hazarded.

Monopolistic and Monopsonistic Interferences with Resource Use

In a purely competitive economy, resources would tend to be allocated roughly in accordance with consumer preferences as expressed in "dollar votes" in the markets. This is true despite a number of flaws in the processes of adjustment. But in other types of market situation this tendency does not appear. When a firm is a simple monopoly or monopsony or when a group of firms are oligopolists or oligopsonists, whatever the basis on which their positions are established, their search for maximum profits will lead to curtailing of production short of the level that would most effectively satisfy consumer preferences. In many

cases it will also lead to waste of relatively fixed agents like machines and buildings that are kept idle though allocated to the industry.

Short-run allocation of productive agents

Let us suppose that an economy is purely competitive throughout with the exception of one industry that is solely in the control of one firm or that is controlled by a small oligopoly group. In the short run there is no possibility of entry of new firms. Variable costs to the firm or firms in this one monopolistic industry would reflect reasonably accurately the worth to consumers of the services of marginal units of these productive services in other employments. Consumer demand for the product of the monopoly or oligopoly industry would serve as a reasonably good index of the marginal worth to consumers of various quantities of products of the industry. Consumer preferences would be realized if in this industry employment of each type of productive service was carried to a point at which average expenditure and marginal sale value were equal. But this will not be the tendency of entrepreneurs in a monopoly or oligopoly situation, who are seeking to maximize profits. They will curtail production short of this point in order to maintain higher prices. Whether the curtailment is in a single simple monopoly firm, whether it results from nonaggressive price competition of oligopolists or from price maintenance or market-sharing agreements among them, the result is the same. The output of the industry is less, the employment of resources in it is less than would consistently reflect consumer dollar votes. More of these resources must find employment in the rest of the economy. As between the monopolized industry and the rest of the economy there is a malallocation of resources; "too few" are employed in the monopolized industry, "too many" being left for other uses.

If it now be assumed—a more realistic assumption—that a large number of industries are monopolistic, this conclusion takes on greater importance. Curtailment of production in the monopolistic areas may in the aggregate become serious, and the forced diversion of resources into less productive employments in purely competitive spheres is a real loss from the point of view of the society as a whole. Within the competitive sectors allocation from industry to industry will tend to be consistent with consumer preferences; within monopolistic sectors it will be erratic; between monopolistic and competitive spheres it will be clearly out of line with consumer preferences, curtailment of use of resources in the monopolistic spheres resulting in an overflow into competitive industries.

An analogous curtailment of resource use occurs in monopsony and oligopsony situations, as a result of the manipulation of the market to hold down the prices paid for productive services. Again, whether the curtailment in inputs is in a single simple monopsony firm, whether it results from nonaggressive competition in buying by oligopsonists, or from a price maintenance market-sharing scheme in the purchase of productive services, the result is the same. The output of the industry is less, the employment of resources in it is less, than would be consistent with consumer preferences as expressed in dollar votes. Unable to find employ-

ment here, more of the owners of productive agents seek outlets in competitive sectors of the economy, and a malallocation of resources as between the monopsonistic and the competitive spheres results.

Long-run under- and overdevelopment of an industry

Short-run allocation of resources is a matter of the channeling of relatively variable agents into firms in one industry or in another. In the long run, adjustments include the shifting of firms and of the relatively fixed agents that they employ as well as the shifting of relatively variable agents. In a purely competitive economy all agents would tend to be allocated roughly in accord with consumer dollar votes but in an economy such as has existed in the United States this is not the case. Long-run adjustments may lead either to under- or overallocation of relatively fixed agents to monopolistic spheres; though in instances of excessive fixed agents in the monopolistic spheres these agents are underused and lie at least partially idle.

1. Where insiders practice methods of exclusion of new firms—misuse of patent rights, manipulation of banking and transportation privileges, unfair competitive tactics, and so on—the obvious result is the underuse of relatively fixed as well as variable agents in this industry or production area. Resources excluded from this or these sectors of economic activity are forced to seek employment elsewhere, where the sale values of their marginal products are less. This is an obvious interference with the realization of consumers' preferences in the allocation of resources.

2. Monopolistic and monopsonistic firms or groups of firms that are unable to exclude new enterprises successfully may find that in the long run the industry tends to become "overdeveloped" through the appearance of these firms. The maintenance of high selling prices and low purchase prices as a means to greater profits results also in the attracting of new firms into the industry even though existing firms could more economically expand production to turn out the same aggregate output. As these new firms become established, making investments in relatively fixed agents, the share of the market held by each firm becomes smaller and smaller; larger and larger proportions of the equipment in the industry is then underused, with frequently very great waste as viewed from a broad social standpoint. If an industry started out as a monopoly or monopsony this process would convert it into oligopoly or oligopsony; but the group of firms would still follow policies essentially like those of the single firm in the monopoly or monopsony situation. The excess of relatively fixed agents would be underused, output of the industry would be restricted, and the allocation of relatively variable agents to the industry (or production area) would be "too small."

The Relation Between Payments to Productive Agents and "Economic Contribution"

It is frequently regarded as a corollary of the philosophy of individualism that owners of productive agents "should" receive for the services of

these agents just the "economic contribution" that they make. But "economic contribution" unless further defined is extremely vague in meaning. In their study of economic processes, economists have arrived at a commonly accepted interpretation of what it means; according to their usage, the "economic contribution" made by a productive agent is the sale value of its marginal product. The reason they identify economic contribution with this marginal sale value instead of with marginal revenue is obvious after an examination of the situation. Suppose that by withholding some of his services a man can get a bigger total money income, so that additional offerings would actually result in a negative marginal revenue; is he making a greater economic contribution when he produces less? Most people would say "of course not," and add that the question was very silly. Yet people do sometimes get bigger incomes in this way than by producing more. Chapters 11 and 20 discussed the fact that the sale value of marginal product provides a rough index of the worth to consumers of the services of a marginal agent in a particular firm or industry under purely competitive conditions. It is a yet rougher index when the economy is riddled with monopolistic and monopsonistic situations; but it serves this role better than any other measure. Under what conditions do people in fact receive for their labor services and for the services of productive agents that they own payments equal to the sale value of the marginal product resulting from use of these services?

Payments and "economic contributions" in a purely competitive economy

Under conditions of pure competition each entrepreneur seeking to maximize profits hires productive agents so long as average expenditure is less than or equal to marginal sale value. This average expenditure is the wages of particular kinds of labor, the price or rent of particular kinds of machines. Thus in each firm inputs are adjusted in such a way that marginal sale values are approximately the same as the return to the owner of the productive agent. Through the simultaneous actions of the aggregate of firms purchasing productive services and of individuals or firms selling these services, payments are bid up or down until (1) the amount of a service offered and the amount wanted are equal, (2) productive agents are contributing in each firm marginal products the sale values of which are approximately equal to the payments they receive, and (3) no shift of employments would yield either bigger payments or greater sale values of marginal products. In a purely competitive economy payments received would match very closely the "economic contributions" made by the agents involved.

Monopolistic enterprise and payments to owners of productive agents

Wherever a firm has some degree of control over price, whether in a situation characterized by simple monopoly, by oligopoly, or by monopolistic competition, marginal revenues will be less than the sale values

of marginal products, though in most cases of monopolistic competition the difference is slight. Entrepreneurs of such enterprises will curtail production, and hence inputs, short of the point at which average expenditures would equal these sale values. As a result agents employed in monopolistic enterprises will be receiving less than the marginal contributions that they make as viewed by consumers; payments to owners of these agents will be less than the "economic contributions" resulting from the services rendered. This is illustrated graphically in Figures 12—1C and 12—1D. If there are many sectors of the economy in which there is such curtailment of inputs, these resources seeking employment in the competitive sectors will add less and less to products there; so their incomes in those sectors will therefore be less, and the "economic contribution" that they are able to make will be less than would have been possible in the absence of monopolistic restrictions.

Monopsonistic markets and payments to owners of productive agents

Monopsonistic situations also involve curtailment of inputs and the payment of incomes less than the "economic contributions" of the productive agents involved. In this case the curtailment is due to the opportunity of the entrepreneur to hold down payments by restricting inputs. Instead of adjusting according to a comparison of *average expenditures* with marginal sales values he focuses on *marginal costs*. The greater the difference between marginal costs and average expenditures the greater will be the discrepancy between the payments received by the owners of productive agents and the value of their services. This is true whether the firm sells competitively or monopolistically: It is illustrated graphically in Figures 12—3B and 12—3D. Again, curtailments of inputs in monopsonistic sectors of the economy forces more agents to seek employment in competitive sectors and pushes down the marginal sale values of their contributions there.

The Effects of Quality Competition

Thus far only slight attention has been paid to quality competition and its effects on economic adjustments. It is, of course, closely associated in practice with policies directed toward "creating" demand; but to understand its effects it is necessary to distinguish at least conceptually between quality competition in response to consumer preferences and that associated with policies intended to influence and change these preferences. In this section attention will therefore be turned briefly to the effects of competition *in response to any given consumers' preferences as between different physical products and conditions of sale.*

The significance of shifting emphasis from price to quality competition

It is quite generally agreed that there has been an increasing shift in emphasis from price to quality competition in recent decades in the United States. What is the significance of this shift? The general wel-

fare would probably be best served by a reasonable emphasis on quality—to prevent shoddy work and inefficient products from appearing on the market—but at the same time by an emphasis on price to make available to consumers products comfortably within their income ranges wherever possible. Have these conditions been fulfilled? No complete answer to this question is possible, but certain general characteristics of quality versus price competition provide a clue.

1. Increased emphasis on quality competition may, and often does, lead to greater efforts in technological research for the improvement of products. This tends to bring better products to the markets and at lower costs. Where, on the other hand, there is aggressive price competition between sellers of slightly differentiated products, there may be a tendency to skimp on quality. This is especially important where inferiority is not readily discernible, as in many consumers' goods.

2. But the increased emphasis on quality competition may also block response to consumer preferences by diverting competitive behavior away from price competition.

3. Emphasis on quality is sometimes misused by a group of dominant producers who get together to establish extravagant minimum standards in an industry, ostensibly as a protection to consumers but actually as a means of excluding legitimate competition from lower qualities. Efforts of leading firms in the plumbing industry to prevent the sale of "culls" or "seconds," and efforts of leaders in the bedding industry to prevent the use of secondhand materials are clear examples.

Caps and wasteful duplication

Quality competition frequently results in the clustering of offerings in a narrow range with very minor aspects of differentiation, while large gaps in offerings may remain unfilled. This is due partly to lack of imagination and the desire to avoid the risks involved in innovations. It is due partly to the advantages of coming in just "on the edge" of what other firms are doing. An example will serve to show the importance of some of these defects in the system.

Europeans commonly produce and use very small cars that have gained no foothold in American markets. For a number of reasons entrepreneurs refrain for the most part from offering such cars. Consumers are now used to the heavier cars and would shift only slowly if at all to the lower-priced models. Prestige has become an increasingly important factor in car ownership, and a manufacturer entering a low-price field would share very little of the general gains in sales from the big advertising campaigns put on by producers of the bigger cars. His chances of making profits are slim and uncertain. Moreover, if such a manufacturer began to have any success he could expect that the other car manufacturers would fight him with every technique at their disposal because he would threaten their basic positions; such a possibility would give him pause. By offering a car very similar to the others he could share in an already well-

established market with much less fear of having to battle an entire group of powerful concerns. He is therefore likely to confine his quality competition to minor details such as chromium trim, streamlined fenders, and free wheeling, while whole ranges of production possibilities go unexplored.

Paying for variety

Sellers seeking to obtain profits by producing goods in the forms and under the conditions of sale that will attract customers differentiate their offerings one from another. In some spheres of production an almost infinite variation of product is possible. This product differentiation arising out of quality competition may be viewed as a more refined aspect of the providing of a range of clearly different classes of goods—shoes, gloves, hats, meat, vegetables, automobiles, radios. When refined quality competition in consumers' goods is taken into account, pure competition appears no longer as an alternative to monopolistic competition, since monopolistic elements arise inevitably to some extent as a part of the process of providing a detailed variety in goods and services in response to consumer demands. There is no doubt that many consumers are willing to pay something in higher prices in order to have the variety made possible by many instances of product differentiation.

Variety costs consumers something for several reasons. Insofar as one seller is differentiated from others by his particular variation of offerings, he obtains at least a slight degree of price control. He will typically use this control to charge prices slightly higher than would result if he were to carry production up to the point at which marginal costs equaled marginal sale value; but unless his product is very markedly differentiated from others this effect on price will be very small. Associated with the differentiation of product is frequently greater marketing costs. The processes of buying and selling in intermediate stages between factory and consumer is considerably complicated. These greater marketing costs are ultimately reflected in the prices consumers pay.

How much are consumers willing to pay for the sake of having a variety of products from which to choose? The answer to this question is difficult even in abstract analysis. For a number of reasons the degree of variety that would arise if differentiated sellers competed aggressively in both price and quality is regarded by some as a sort of "consumers' ideal" degree of variety (assuming also that consumers are fully informed); it might equally logically be regarded as a maximum of variety in relation to consumers' preferences, as the maximum degree of variety for which consumers are willing to pay. Although the assumptions are somewhat arbitrary, by accepting this criterion for the moment two general conclusions may be drawn: (1) Under conditions of monopolistic competition the degree of variety that would develop would be the maximum for which consumers would be willing to pay had they an opportunity to express their preferences for variety versus lower price. (2) Heterogeneous oligopoly with free entry to the general production area

(though not to duplication of the offerings of any single firm) results in an overmultiplication of varieties. "Too many" firms appear, each firm operating at an inefficiently small output level with a small share in the total market for the group of related products; the variety is greater than would result had aggressive price and quality competition cut down profits before so many firms appeared.

Errors in sellers' judgments of consumer demands

The differentiation of products involves a large amount of guesswork concerning the various elements in a product and the conditions of its sale that are in fact attracting customers. It is difficult for sellers to judge which services and which aspects of product differentiation are really worth to consumers what they add to the costs of production. This is especially true of complex goods such as automobiles. Very commonly consumers pay for some unwanted gadget that "comes with" the car which they select for other reasons. Much has been written about costs of credit and delivery, which are commonly spread over customers not using these services as well as those using them. In some cases an attempt is made to segregate such costs and charge only those using the services or the gadgets. Unless this is done—and in many cases it is impractical—there is possibility of a serious flaw in the channel of communication between consumer and producer with respect to quality and service preferences.

The Effects of Demand Creation

Most of quality competition is more than response to any given set of consumer preferences; it is associated with efforts to influence and channelize these preferences. Policies of demand creation were examined briefly in Chapter 24, but some of the broader effects of advertising policy on adjustments in the economy will now be more fully considered, under the following headings: (1) the effects of advertising on consumer selections in the market, (2) advertising as a basis of monopoly power, (3) shifting allocation of resources between production and selling costs, and (4) advertising as a way of meeting uncertainties in modern markets.

The effects of advertising on consumer selections in the market

Advertising may inform, it may misinform, and it may persuade. In the first instance, existing underlying preferences become more effectively expressed in the demands for the products of the various sellers; in the second, they will be less effectively expressed; in the third, the underlying preferences themselves may be altered.

1. Insofar as advertising provides information it leads consumers' selections in the markets to be more nearly in accord with their underlying preferences. When consumers are better informed, producers of inferior goods may have the choice of lowering price, raising quality, or being eliminated from the market. They will provide the higher quality high-price good or the lower quality low-price good according to their esti-

mates of consumer demands. Thus informative advertising may increase the effectiveness of response to real consumer preferences since these preferences will be more accurately reflected in demand. Even if all advertising were informative, however, there would be a question as to how much use of resources in providing such information was justified. This is one of those immeasurables which is difficult for even each individual to decide, let alone some economist presuming to make a decision for all individuals. Efforts to obtain legislation providing for information to consumers and for protection against misinformation is a recognition of the fact that collectively people are willing to spend something for the benefits of becoming better informed.

2. Unfortunately, advertisers are not concerned primarily with providing information for consumers. Instead they are concerned with attracting customers to buy their particular products. Wherever such advertising is misleading or out-and-out falsification, it leads consumers to make market selections less consistent with their underlying preferences than would otherwise be the case. They may buy the "wrong" things; they may spend more than necessary on the "right" ones. Ironically, as a tool for creating illusory product differentiation, advertising permits not only the charging of higher prices, but the use of high prices themselves to help create the illusion of quality superiority.

3. That advertising affects people's ideas of what satisfactions are important there can be little doubt. How great these effects are is more open to question. And whether any particular change will be regarded as "desirable" or not will depend on the particular values of the individual or the group passing judgment. Advertising that leads consumers to smoke more cigarets may be regarded as undesirable by some but not by others. Advertising that leads to increased milk consumption may be quite widely regarded as beneficial.

A function of advertising that is commonly approved is its contribution to the informing of consumers concerning new products.¹ Here it clearly expands the horizons of the consumer, both increasing information and changing ideas as to what satisfactions are important. Examples of new products advertised within this generation are rayon, automobiles, radios. Associated with such informational services is the fact that advertising under these circumstances may hasten the adoption of new products by more and more consumers, until they can be made on a mass production basis and thus made available to consumers on lower and lower income levels.

Advertising as a basis of monopolistic power

Typically, advertising separates a seller more distinctly from his rivals, at least temporarily. It is through such differentiation, whether based on

¹ Informative advertising may sometimes be disapproved, when it is believed that particular kinds of knowledge have bad effects. An example is the opposition of some people to the spreading of birth control information.

real differences or on illusions, that the advertiser is likely to obtain the greatest opportunities for profits. Whether the result is monopolistic competition with many producers of slightly differentiated products or heterogeneous oligopoly with a few rival producers watching each other closely, the exaggerated differentiation results in greater opportunities for each firm to boost prices by small restrictions of outputs. Moreover, advertising can sometimes set a seller off from rivals so distinctly that he becomes almost a simple monopolist. For a long time the producers of Bayer aspirin were in this enviable position; even though other producers of aspirin undercut them sharply they were unsuccessful in drawing customers away. While this is not typical it is also not a unique case. Sellers who thus establish themselves may for long periods obtain large profits from the expenditures of the customers whom they have attracted.

When the increased differentiation of products caused by advertising activities occurs through spreading accurate information as to differences in quality it may be that consumers prefer the higher prices plus the greater variety if these are the only alternative to the lack of information and variety that might otherwise prevail. There is no way of measuring these preferences. But where differentiation arises through the creation of an illusion of differences in the minds of consumers, it is clear that the consumers lose all around.

*Shifting allocation of resources between production and advertising

The cumulative piling up of competitive advertising has the effect of a cumulative diversion of productive resources from the production of goods and services in response to any given set of consumer demands to efforts to influence these demands. Insofar as the efforts of one seller are simply neutralized by those of another seller, or simply shift production from one firm to another offering an essentially identical product, there is obviously a waste of resources, since society has no net return from these efforts. The selling costs are added to the costs of production, and result in higher prices. Taking all producers together, there will be a smaller total output of goods and services than could otherwise be obtained from the same resources.

Advertising as a way of meeting uncertainties in modern markets

Businessmen produce to fill an estimated future demand. When goods appear on the market the producer must find buyers for them. He therefore proceeds to use whatever means are at his disposal to insure that consumers will buy what he has to sell. Though some of the uncertainties of production are created by advertising itself, it is also true that advertising may be a means of protecting industry and the general public against wastage in production of goods where demand for different varieties is otherwise very erratic and unpredictable.

Product Diversification and the Character of Competition

Product diversification has many important economic effects. Many of the causes listed in Chapter 5 as giving rise to such diversification

indicate that it may be important in reducing production costs and cutting out unnecessary waste. It may also have an important effect on the market position of the concern as a buyer and as a seller. And it may give greater stability to particular business enterprises. Here attention will be focused on the effects of product diversification on the character of competition, including the extent to which monopolistic situations are encouraged or discouraged.

Response to consumers' preferences in joint-product industries approximating the model of pure competition

If an industry were purely competitive, and producers were fully informed concerning alternative opportunities, the results would tend to be allocation of resources in accordance with consumers' preferences. As was indicated in Chapters 18 through 20, this would be just as true of industries characterized by concerns producing several products as by concerns producing one product only. But the problem of cost estimation is made much more complex, and this leads to some blurring of the picture. Inaccurate estimates of cost mean inaccurate responses to consumers' preferences for the various commodities involved.

The use of some products to increase sales of others

Frequently one or several of the goods produced in an enterprise are sold at low prices in order to attract customers to purchases of other products on which the concern anticipates making most of its profits. Examples are manifold. Utility companies may sometimes sell some kinds of electric equipment at very low prices in order to increase the consumption of electricity. Oil lamps were actually given away in China by oil companies seeking to sell more kerosene.

Long and loud have been the complaints of manufacturers of branded goods and of small grocers against the offering of "loss-leaders" designed to attract customers into retail stores; this is simply another case of selling some products either below cost or at very low mark-ups as an advertising device. Frequently it involves deception in giving the impression that in general prices in a particular store are lower than elsewhere when in fact this is not the case; on the other hand, many complaints arguing that there were such practices are just the shouting of inefficient producers against legitimate price competition from more efficient low-cost rivals.

Product diversification as a basis of strength in conducting tactics of exclusion

Efforts to gain and to maintain dominant positions in an industry are frequently very expensive. So also are some policies aimed at maintaining a market position for the future even at a temporary loss. Where such temporary sacrifices appear to be worth the cost, a concern producing a diversified line will be in a relatively strong position since these activities in any one product may be carried for a while by profits on other products. Two examples will serve to clarify this point.

One of the techniques by which a concern or group of concerns gain and hold dominant positions in the market for a product is the use of extreme price-cutting tactics to drive out potential or budding rivals. A large national concern producing one product only may carry such price wars locally through the profits made elsewhere; but sometimes this is not easily done. Product diversification provides an alternative. Price-cutting to attain or maintain monopoly position with regard to one product may be supported through the incomes on the other products.

Sometimes when a market is especially depressed a concern continues to maintain prices at previous levels despite severe curtailment of output which for the time being involves great sacrifice. It maintains these prices in order to avoid "spoiling the market" with respect to future possibilities by cutting prices now. Such a policy on a particular commodity is much more easily carried out if the firm has other production lines to tide it over; this is especially true if equipment can be shifted in part from production of the price-maintained good to other products and back again.

Product diversification and product differentiation

The more effectively a concern can differentiate its product from others in the minds of buyers, the stronger will be its position in obtaining profits. When a brand name becomes familiar to the public it carries prestige for the various different commodities of the same concern. The advertising of one line encourages sales in other lines as well. General Motors' products mutually support each other in this way, so do the products of each of the big cosmetics producers, and of well-known clothing stores. The advantages of a firm advertising a full line of products is probably greatest where consumers' goods are involved, whether on the retail level or in wholesaling and manufacturing.

Where products of different firms are but slightly differentiated, the profits of one concern are continuously checked by competition of rival producers of close substitutes. This is specially true in markets for finished goods. When a firm extends its control over the production of competing products, whether by outright purchase of previously rival companies or in some other way, an enterprise or monopolistic group is in a much stronger profit position. Consumers are no longer free to choose substitute products sold by other firms if they consider the prices charged by one firm to be too high or the quality of its products to be inferior.

On the other hand, under certain circumstances diversification of products may increase effective competition. When producers of well-known brands have attained control over a market for a particular product, say washing machines, it is very difficult for a new and unknown concern to make any headway. But a large concern such as General Motors already well known for other products might add washing machines and succeed in cracking a previously dominant group in the washing-machine industry, insofar as dominance had rested on the use of advertising and trade-marks.

Entrepreneurial Mobility and Economic Opportunity

In a dynamic society there are continual changes at one point or another that start off a long series of interconnected readjustments. The more mobile are the productive resources of the economy the more readily will adjustments to change occur. One of the factors in this process of adaptation is the withdrawal of firms from unprofitable activities and the establishment of new firms in expanding and profitable sectors of the economy. Anything that checks this adjustment of entrepreneurial activity hampers the response of the economic system to changing consumer preferences; it also interferes with the realization of income opportunities potentially available to the immobile or sluggish entrepreneurs. The question of the extent of entrepreneurial mobility and of impediments to that mobility is therefore very significant in the evaluation of the adjustments that take place in an economic system.

Do entrepreneurs move freely from one form of enterprise, one industry or production area, into another? There is no yes or no answer to this question. It depends on the individual entrepreneurs—on the extent of their knowledge of alternative opportunities, on their preferences, on individual difficulties of readjustment, and on the degrees of conservatism or restlessness of these individuals. It also depends on the degree of specialized knowledge necessary to function successfully in an entrepreneurial capacity in firms in different industries, on how large an investment is necessary to start in business, and on the presence or absence of artificial restrictions of entry to particular industries or production areas. And it depends on the extent to which entrepreneurial functions are divided and performed by different people.

Ignorance of alternatives

The making of the basic policy decisions in an enterprise is frequently a very complex problem. It requires a broad knowledge of the markets in which productive services (labor, machines, raw materials, and so on) are purchased, of the techniques of production in the industry, and of markets for the sale of the product. Moreover, an understanding of currently existing conditions is not sufficient for successful entrepreneurship. Entrepreneurial decisions are based on anticipations of future revenues and costs, of future demands of customers and offerings of productive services. It may not be difficult to see that firms in a certain industry are currently making profits and that this looks like a good possibility at the moment; but it may be much more difficult to judge accurately whether a new firm by the time it is operating will still be a profitable venture. Lack of full knowledge of alternative opportunities is almost universal.

That inadequate information can act as either a check on mobility or a factor exaggerating such mobility of entrepreneurs into new lines of production is evidenced in many spheres of economic activity. Without detailed knowledge of an industry some men may hesitate to leave one

line of activity to move into the less familiar area unless present activities are bringing very unsatisfactory results and the risks of change are small. Others may welcome the adventure of entering into new and untried realms of action. And in many cases entrepreneurs move into an industry which is in fact quite unprofitable under the illusion that the pastures on the other side of the fence are greener.

Entrepreneurial preferences

Some types of entrepreneurship are rather sharply marked off from other spheres in the kinds of activity involved and the way of life associated with the business. Preferences for farming may, for example, discourage some people from shifting out of farming into manufacturing or merchandising; though shifts between different types of farming may be undertaken. Preferences for the gambling aspects of "the oil game" in Texas may hold men there despite more profitable alternatives elsewhere, and vice versa. Preferences for living in one place rather than another may check some men from shifting into entrepreneurship in industries in which business could be carried on successfully only by moving to a less preferred location.

Costs and difficulties of readjustment

"Learning the ropes" in a new industry may be a difficult, a time consuming, and for a while a costly, undertaking. The complexity of the role of entrepreneurship, which may contribute to relative ignorance concerning the possibility of profitable alternatives, is also a check on entrepreneurial mobility from industry to industry. Making the change requires too much effort and for a time too much money. To shift from farming to manufacturing automobiles or women's clothing would be quite a jerk; though to shift from farming to running a machine-shop or a slaughtering plant might for some individuals be quite feasible. Where functions of making management decisions are more sharply separated from investment, and are subdivided along special lines—purchasing, technical production management, selling—it is, however, not so much entrepreneurial immobility that hinders in the setting up of a new firm; it is rather the degree of mobility of salaried professional men from one industry to another that becomes a matter of importance. While the entrepreneur who is simultaneously investor, manager, and broad policy-maker may be checked in shifting his line of activity because of the difficulties and costs of learning the new business, this is not so true of large investors in corporate enterprises who study market opportunities in determining where to invest but who hire specialized managers to conduct the business within the broad policy program which they lay out.

Size of initial investment

That the amount of investment necessary to start a business is an important factor in determining the degree of mobility or immobility of firms (entrepreneurs) into an industry is evidenced in data on the number

of new enterprises and the mortality rate in different lines of activity. Retailing affords an example of an area in which initial investment is small and profit expectations of new firms are very low; the high mortality of firms in retailing can be explained only by recognizing the extent of ignorance of those entering this field coupled with unfavorable alternative opportunities for many of these individuals and the very small initial outlay necessary to get started. At the other extreme, the necessity for large initial investments explains *in part* the slow growth of new enterprises to compete with the United States Steel Company and the fact that for years no competitor arose to challenge the Aluminum Company of America, despite the high profits being made in these industries.

Restrictions on entry

Cost of initiating new enterprises is only a partial explanation of the slow development of competition in steel or oil-refining or like industries, or of the failure of any firm to appear beside the Aluminum Company of America. Techniques of restricting entry to particular industries or production areas have already been indicated briefly in discussions of monopoly and oligopoly. They will be considered in detail in later chapters. That they are extremely important in many instances is beyond doubt. Such restrictions differ from other checks on entrepreneurial mobility in an important respect. They constitute direct interferences by "insiders" with the free opportunity of outsiders to engage in whatever lines of productive activity they may choose. These restrictions are therefore directly in conflict with one of the basic tenets of the philosophy of individualism in American ideology, despite the fact that they are practiced by men who most loudly acclaim the virtues of that philosophy. Moreover, along with most of the other factors that check entrepreneurial mobility, they impede the channeling of productive resources into those uses that would reflect most effectively consumer preferences.

Labor Mobility and Economic Opportunity

Probably more important than checks on entrepreneurial mobility are the limitations on the mobility of labor from firm to firm, occupation to occupation, place to place. And associated with this labor immobility are some serious interferences with equality of opportunity of different individuals to qualify themselves for lucrative employments and to find such employments. Most of the factors that impede labor mobility and that block opportunities for employment in line with potential contributions may be classified under one of the following headings: (1) ignorance, (2) preferences, (3) differences in ability, (4) costs of shifting employments, (5) discrimination against particular groups of people, and (6) exclusive tactics of organized groups of laborers. Each of these groups of factors will be examined briefly, and a general evaluation of the allocation of labor services and the economic opportunity open to individuals in the sale of labor services will then be undertaken.

Ignorance

While the quantitative importance of ignorance as a factor impeding entrepreneurial mobility may be open to question there is no doubt that it is extremely significant in many labor markets. The ignorance of alternative opportunities in firms of the same kind in the same locality causes some lags in adjustments, but this type of ignorance is probably relatively insignificant. More serious is the lack of information concerning opportunities in alternative occupations, and in other places. For example, there have long been opportunities for women in a number of kinds of work using home economics training, with only a slow awakening among women to the existence of some of these possibilities. Similarly there are lags in discovering new job opportunities for men arising out of the development of new technological innovations. And such ignorance concerning other occupations is by no means confined to lack of knowledge of opportunities in newly developing fields. Moreover, ignorance of opportunity in other places is equally important, especially the ignorance of some rural people concerning opportunities in urban industry and the ignorance of eastern urbanites concerning opportunities in the interior of the country. Increasingly smooth operation of vocational bureaus and employment exchanges is removing some of this ignorance of job opportunities, but the gap in knowledge is still great.

Preferences

Preferences may again be occupational or geographic, and they may also involve preferences (or pressures) for a larger income now at the sacrifice of future income.

The effects of preferences for one job over another may be conveniently divided into two categories: (1) those preferences that vary between individuals because of special likings for particular types of work, and (2) those preferences that are common to most workers or large groups. Preferences of both types diminish the mobility of workers from preferred occupations, increase their mobility to these positions; but they operate in different ways. Where preferences are individualized, there will usually be enough workers who are indifferent to the particular factors involved so that over a long period the preferences will not be a cause of persisting differences in wages (no other impediments to entering the occupations involved being assumed). Where preferences are shared by large groups of workers (as the general preference for white collar over manual jobs), however, differences in wages between occupations of comparable skill may persist.

One of the most important factors checking geographic mobility is doubtless the preferences of individuals for the localities to which they are accustomed. Here are their friends, here is the kind of life that has become a basic part of their satisfactions, here are the landscapes that have become dear to them. Great differences in possible planes of living

may be insufficient to draw them away from these things, which have become so important to them.

Time preference in the labor market is frequently a very important consideration. The desire to have a larger income now frequently leads workers to choose those jobs bringing the greatest immediate returns, even though they may be "dead-end" jobs that yield much smaller average incomes over a life time. In some cases immediate pressures may seem so severe that long-run considerations are out of the question, in other cases it is simply a matter of foreshortened vision. This short-run approach to job opportunities finds its most extreme expression in some cases of child labor that cut off the years of education, and in the choice of occupations requiring a high degree of manual speed and dexterity. It is partly because of these time preferences among workers that some firms are able to drain their workers by pressures that markedly shorten the lengths of their working lives.

Differences in abilities

There are obviously differences in ability between different individuals. If there were complete equality of opportunity for all to obtain training for any kind of work preferred, there would still be differences in abilities between people that would persist. These differences would in fact reflect both "natural" differences in "intellectual" capacity and manual dexterity, and differences in determination and ambition. Some people may be generally more capable than others. Some may have special talents of one kind, others of another. Such differences are among the "givens" in the potential human resources of a country; they are one of the reasons why specialization of occupations is economically advantageous; and they are also one of the reasons for persisting differences in wages between different occupational groups. They obviously limit the degree of mobility possible in the shifting of workers from one occupation to another.

Costs of shifting employments

The costs of shifting from one occupation to another or from one place to another are sometimes prohibitive, and frequently seriously interfere with labor mobility.

Costs of training for some occupations are much greater than for others. Such costs would not usually be incurred unless higher wages were anticipated as a result, although occupational preferences may lead some individuals to incur expensive training despite poor prospects of recouping these additional expenditures through later differential incomes. Actually such costs discourage entry into certain occupations where rewards would more than compensate, because many people are unable to meet them in advance. It is very difficult, for example, for an individual with no economic backing to finance himself through the period of training required to become a physician.

That it costs money to move and that there are risks in moving unless a

job in the new community is definitely assured are too obvious to require much comment. Costs of moving include not only the sums paid for freight and passenger transportation. There are also, for example, sacrifices involved in selling a house and buying another one, and in many other necessary re-adjustments. Where a definite job is not already arranged, costs to the individual may include a waiting period while job-hunting in the new locality.

Discrimination against particular groups of people

One of the most inconsistent aspects of social structure in the United States is the existence side by side of a democratic philosophy and extreme discrimination against particular minority groups. It is a well-known fact that the occupational opportunities open to Negroes in all parts of the United States are very much restricted by social customs and the attitudes of the people, even though they supposedly stand on equal terms in the law. Some other racial minority groups are also underprivileged in this respect. These differentials work against the Mexicans of the Southwest, for example; and the relative lack of opportunities available to the American Indians has shocked many a reformer. There is probably no type of stratification of opportunity more rigid than that that follows race lines. It would be folly to attempt to predict trends with regard to persecution of racial minorities in the social unrest of contemporary life; but some rough indication of the current disadvantages of Negroes is indicated in Figure 35—5 of Chapter 35 to follow. Such restrictions limit the offerings of certain labor services by excluding a potential part of the supply.

Somewhat analogous to the discrimination against particular groups in occupational opportunities are restrictions on the migration of certain groups. Restrictions on immigration into a country are an obvious example. Less obvious, but of considerable importance is the pressure on low income families to remain where they have a legal residence and are therefore reasonably sure of being able to obtain government aid.

Exclusive tactics of organized labor groups

The role of trade-unions in the labor markets of the United States will be the subject of some later chapters. A discussion of restrictions on mobility of labor would, however, be incomplete if the exclusive practices of organized labor groups were ignored. Largely through big initiation fees and exaggerated apprenticeship rules, coupled with agreements with employers that only union workers will be hired, some trade-unions effectively limit the number of men entering a trade. These limitations may extend over the entire country, or they may be primarily local; but wherever they exist they clearly interfere with freedom of economic opportunity to the "outsiders." Thus far it is in only a few occupations that such trade-union controls have become highly significant; for the most part trade-unions are not effective devices for excluding workers from employment in the occupations in which workers have been organized.

Labor mobility, resource allocation, and economic opportunity

Insofar as labor immobility is due to "innate" differences in abilities, it is inevitable and therefore necessarily accepted as part of the framework within which an economy operates. But all other types of limitation on labor mobility act as interferences with the allocation of labor resources in accordance with consumer preferences. Immobility resulting from the preferences of the workers themselves has been mentioned before as interfering with the realization of consumer preferences in resource use; but since freedom to choose the job of one's preference is commonly regarded as an important aspect of a "desirable" economic order, it may be largely dismissed from present discussion with this passing comment.

There remain ignorance, costs of shifting employments (and of acquiring training in some employments even initially), discrimination against minority groups, and direct exclusions from certain occupations and certain places. All of these factors lead to malallocation of labor resources, and in two ways. They may sometimes place a firm in a monopsony position, or provide the basis for oligopsony control of labor markets with resulting restrictions on the employment of workers in these sectors of the economy, coupled with the depressing of wages. And they directly slow up and sometimes block completely the shifting of labor resources from less to more productive employments. This failure to make the most effective use of potential labor resources is a blot on the scutcheon of a supposedly democratic country, not only because of the sacrifice in national planes of living, but also because of the extent to which privileged groups maintain for themselves high incomes at the expense of other people whose opportunities are limited by circumstance and by directly exclusive policies.

Some industries are "parasites" on labor. Workers in these industries actually deteriorate both physically and mentally because of the low wages, and the bad working conditions (including danger of accident) to which, through the pressure of necessity and through ignorance and short-sightedness, they are subjected. The slave owner must count the maintenance of his slaves in health and efficiency as a part of his production costs; but not so the employers of "free" men. The deterioration of workers is both a calamity for the individuals involved and a loss to the society as a whole. These are social costs that are not fully included in the costs incurred by the enterprises in the industry.²

The many possibilities of public policy that might be directed to the curing or modifying of some of these faults in the functioning of the system cannot be considered here. But four obvious policies may at least be mentioned: (1) the extending of educational opportunities in training to whomever may be interested and capable of learning, for whatever occupations they may prefer, regardless of the economic backing that they may have in their family situations, (2) the development of employment

² That such discrepancies between private and social costs lead to malallocation of resources was pointed out in Chapter 20. Such parasiting in labor markets is an especially serious form of discrepancy between private and social costs.

exchanges and more detailed informational services concerning job opportunities, (3) a policy that would "crack down" on exclusive tactics of both labor groups and employers, and (4) legislation protecting laborers in conditions of work.

Conclusion

The philosophy of individualism and the "atomistic" view of society

In its most extreme form the philosophy of individualism is based on an "atomistic" view of society, on the assumption that the preferences of individuals are arrived at independently and separately from the attitudes and actions of other people, and that individuals separately adjust to environmental conditions on the basis of these independently determined preferences. Consumers' preferences thus individualized appear as data to which other individuals independently adjust in line with their separately existing ideas and purposes; and economic adjustments are conceived as the directing of resources in response to independently determined individual consumer preferences expressed in market purchases. Entrepreneurs similarly make their individual adjustments to the environment around them separately, and without significantly changing that environment through their individual actions. Each individual, whether as consumer, entrepreneur, owner of productive agents, is an "atom" in the total of the economy. It is the interaction of many atoms that is assumed to be the basis of economic adjustments; each separate atom plays an infinitesimal role in the whole. This philosophy is therefore fully relevant to the evaluating of the American economy only to the extent that the economy is or may be made to be atomistic.

If consumer preferences are arrived at through individual decisions, independently of group pressures and cultural influences and independently of producer influences on consumers' underlying choices, then it makes sense to apply rigorously the individualistic criterion to the functioning of the economy—to inquire to what extent producers respond to these independently determined consumer preferences. Insofar as consumers' preferences are a result of group behavior in general or of producer behavior in particular, the individualist criterion is inapplicable. And insofar as the changing of consumer preferences is regarded as desirable, the goal of individualism is of course partially denied.

Similarly, insofar as the economy approximates the model of pure competition, and entrepreneurs and owners of productive agents are buying and selling in markets in which the individual alone plays an infinitesimal role, the productive system may fit the basic assumptions of an individualistic conception of society. When production of a good becomes centered in a single firm or a group of firms, controls cease to be the automatic results of atomistic adjustments. Monopoly, monopsony, oligopoly, oligopsony—these are nonindividualistic. In their very nature they are contrary to the tenets of individualism in its complete form. Moreover, even an otherwise purely competitive industry, such as wheat farming,

ceases to fit the extreme model of individualism when the individuals involved combine for group political and economic action that supersedes the atomistic determination of economic adjustments. And the organization of laborers in trade-unions is again a substitution of what might be termed "groupism" for "individualism" in labor markets. The increasing importance of group action in the many sectors of modern economic life is not to be regarded as some mysterious aberration; it is rather the logical result of increasingly informed action on the part of individuals who in seeing their common group interests forego something of individual independence in order to profit from coordinated action. Criteria of atomistic individualism are still applicable insofar as production is characterized by atomistic adjustments, and insofar as it is practicable to break down group controls and revive atomistic competition. But where individual action will almost inevitably evolve into group action, to adhere blindly to the more extreme tenets of the individualist philosophy is simply to perform the act of the ostrich that puts its head in the sand. Such developments call for a reconsideration of the values of individualism, and of the possibilities of preserving some of these values, while taking into account the inevitability and the significance of power concentrations with a view to controlling their expression in the interests of a re-formulated conception of the "general welfare."

The strength and ruthlessness of "rugged" individualism

As individualism found expression in the nineteenth-century American economy, it was a philosophy for the strong man, rewarding strength. It was ruthless in its destruction of the weak. Equality was interpreted as equal opportunity to join the struggle. Until recently a policy of minimizing the intervention of government in most economic activity was viewed as the way to attain the maximum of individual liberty in economic affairs. Men were assumed to have equal economic opportunity so long as government played the negative role of doing nothing directly to interfere with opportunity, although there were some exceptions to this position. Americans were to be "go-getters" and it was the "self-made" man who represented the ideal. The expression of this assumed equality as between individuals became in many cases a hard-hitting competitive struggle. Darwin had argued that in nature the fittest would survive; and so it was assumed that a free private enterprise system, left to the play of "natural" forces, would weed out the inefficient and lazy, reward the energetic and able. "Let the best man win"; how frequently has this significant phrase crossed the lips of Americans sharing the philosophy of rugged individualism!

The individual was supposed to be able to take care of himself not only as a producer and income receiver, but as a consumer as well. Associated with the "survival of the fittest" and "let the best man win" was the parallel maxim—"caveat emptor," or "let the buyer beware." The consumer was supposed to be sufficiently alert to protect himself against those who might cheat him.

But equality of opportunity in the theory of rugged individualism is not the same as equality of opportunity in fact; and the most expert of consumers is inadequate to protect himself in the complexities of the modern market. The rugged individualism of a pioneer country is giving way to changing concepts of what is meant by true equality of opportunity in economic affairs. It is not only the inaccuracy of an atomistic interpretation of the functioning of a society that is bringing about new evaluations of what is involved in "general economic welfare." It is also the helplessness of individuals caught in the web of an intricate exchange system such as was unknown a century ago that gives rise to changing ideas among "individualists" as to what expression may be given to individual freedom and to what they regard as fundamental values of the individual in a modern democracy.

The contrasting extreme in a philosophy of totalitarianism

At the opposite pole from nineteenth-century individualism is the philosophy of totalitarianism; and in its extreme form it is equally impractical of application. A totalitarian philosophy idolizes the state; individuals are simply cogs in the machine that is to be operated in the interests of the state. A completely totalitarian economy would be one in which a dictator who speaks for "the state" would tell individuals what they should do as producers, what their incomes should be, and exactly what they should consume. There would be no room left for the expression of individual differences in tastes and preferences. Individuals would simply play the passive roles of robots following mechanically the orders of "the supreme one" whose voice was the voice of the state.

Even in the most complete centralization of control in the Russian communist state and in the German war economy this degree of totalitarianism has been unattained and unattainable. Incentives are necessary to persuade individuals to perform the functions assigned to them, as was the sad lesson learned by Russia when she attempted to clamp down too hard on the peasant farmers. Even with the strong control of a German Gestapo there will be evasions of tightly drawn rules that limit individual freedom to an extreme degree; the Germans have recognized this fact in the intricate system of checking on behavior of individuals and in the extra food rations allowed to some groups. Although the German war economy may approach the extreme of totalitarian control, it is not and cannot become completely totalitarian; just as even in its most individualistic days the American economy had many nonindividualistic elements.

Evolving concepts of individualism in American Democracy

The "humanitarianism" of modern America has turned a searchlight on some of the problems of "underprivileged" people, thus recognizing the fact of inequality of economic opportunity. There is increasing recognition of the limitations of atomistic interpretations of economic life, of the important role played by power and pressure groups in the economy,

and of the intricacy of the system and the great insecurity faced by many groups and individuals. These changes in analysis of the functioning of the system are leading to changes in the formulation of social goals and of policies directed toward attaining these goals. Government is viewed as playing a far more positive role today than formerly. Democracy as expressed in current American ideologies appears to be taking on a more collective character, though in denying the atomistic individualism of the rugged era it is nevertheless retaining emphasis on the importance of the rights of individuals. This new phase of social ideology is expressed in President Roosevelt's statement of "the four freedoms": freedom of speech; freedom of worship; freedom from want; freedom from fear. While these words have not the dramatic greatness of Jefferson's historic statement of rights in the Declaration of Independence, or of Lincoln's speech at Gettysburg, they nevertheless represent the developing interpretation of the meaning of the rights of man as seen in the setting of modern American economic and social organization. It is significant for the formulation of public policy in the economic sphere that emphasis is placed on freedom from want and freedom from fear. American ideology and the American economy itself appear to be moving along a path toward a middle ground. Neither the extreme of individualism nor the extreme of totalitarianism is or ever has been practically attainable; but a particular economy may be nearer to one extreme than to the other. The direction of movement in recent decades in America is clearly further away from the extreme of atomistic individualism and in the direction of increased central planning of the economic life of the country.

Part V

MARKET CONTROL AND PUBLIC POLICY IN AMERICAN ENTERPRISE

PREAMBLE

IF EVERYONE but Peterson and his gang would produce a lot and compete with rivals to get customers by lowering costs and price and improving quality, then Peterson could get rich by getting a corner on a particular industry and producing very little and charging a high price for it. But if everyone could and did act like Peterson, then no one would get rich, not even Peterson himself.

There are a lot of people who would like to act like Peterson, and some of them succeed. Some cannot do it by themselves and they feel left out when they look at people like Peterson. They then get the government to help them out so that they too can get more income by producing less.

But sometimes the government objects to Peterson and his friends. It says "No, you can't do that, it is against the public interest." It may be that the government will not let people like Peterson and his gang fix prices and buy stocks in other companies and divide the market among the members of the gang; it tells them that they have to compete with each other and let other people in on things. Or it may be that the government lets Peterson run a business where no one else competes, as a local water plant, but it tells him what prices he may charge, and a lot of other things that he must do.

People disagree about just what forms government policy toward business practices should take; but anyone who does not get too tangled in the intricate web of rationalization and hair-splitting argumentation can understand these simple truths: (1) that there will be much to consume if everyone produces much and little to consume if everyone produces little, and (2) that we shall not progress very fast if improvements are trampled out whenever they threaten to displace old and entrenched inefficiencies.

How can we get abundant production from our abundant resources? Progress from the work of our progressive inventors and organizers? Here is the nub of the problems we are about to tackle.

CHAPTER 26

Foundations of Power

THE story of American "big business" over the last century is a dramatic one. It is a story of the growth of great business empires, of ruthless cut-throat tactics, and of "gentlemen's agreements," of a shifting pattern of organizational arrangements and business policies to meet changing legal and market conditions, and of a repeated use of political pressures by large business groups. The foundations of the power of these dominant business groups are of many kinds; usually many different bases of control are combined in any single power group. A full survey of these practices and techniques would be quite impossible; but we may examine some of the most important. This will be done under six main headings: (1) Close-knit federations of corporate enterprises, (2) private loose-knit federations, (3) publicly fostered cartelization, (4) controlling and creating bottlenecks, (5) discriminatory price cutting, (6) manipulation of laws concerning standards. There are many overlappings among these classifications, as will become evident in subsequent discussions; but they serve as a starting point.

Close-Knit Federations

The most important types of close-knit federations are the "trusts," "holding companies," and business consolidations. They are all essentially superstructures built on the corporate form of enterprise. There have been two major epochs in the history of such developments, one in the tremendous days of the end of the nineteenth century, the other in the boom of the 1920's.

The corporation in the organizational framework

The great advantage of the corporate form of enterprise is that it facilitates the establishment of large firms in industries in which large enterprise is much more efficient than small; but by the same token the corporate form makes possible the raising of huge sums to build up oligopoly and monopoly positions, to drive or buy out competitors, and to enforce barriers against potential newcomers to the field. Today the gigantic corporation, spreading over the entire country and producing a variety of related products is a commonplace; seventy-five years ago a

billion dollar corporation was a wild dream. Mere bigness does not connote oligopoly or monopoly power, but it is obviously associated with such power. It has been the unrestricted use of corporation financing that has permitted the existence of such huge monopolistic giants as the Whisky, Sugar, Oil, Aluminum, and Steel trusts, to name only a few of the outstanding huge dominant business units of the last half century.

The trusts

The trust device, well known and long used for other purposes, involves the holding of property by a trustee who has power to administer it and receive the income from it, both property and income to be used by the trustee as directed in the trust agreement. This scheme has often been used by persons wishing to have funds held safely for the benefit of their children, and for other similar purposes. The business trust was ordinarily a corporate organization formed to control a number of corporations by making itself trustee of the stock of the component corporate parts of the trust. The original owners of the stock turned their securities over to the trustee corporation, and received "trust certificates" on which dividends were paid. They remained technically owners of the stock but power to vote was transferred to the trustee organization.

A simple example of a business trust may be illustrated as follows: Companies A, B, and C, might be railroads competing with one another; some promoter or group of promoters sets up a trust, to which the controlling stockholders in railroads A, B, and C turn over their shares, receiving in return "trust certificates" showing that the trust company holds stock owned by them. Control of the three competing roads would then be centralized in the hands of the trust, which of course would operate them so as to maximize profits and minimize competition among the three. The earnings of the roads would then be returned to the original stockholders in the form of dividends on the trust certificates. It would be expected that the arrangement would be advantageous both to the promoters, who generally managed to obtain fat fees for their part in setting up the trusts and often retained part control in them, and to the stockholders in the original operating companies, who benefited by the elimination of competition among their companies. Sometimes the trust itself did nothing but hold control of the operating companies, consisting merely of an official office; at other times a large operating company would also act as trustee for the stock to control other operating companies. Of course it was not necessary for the trust to own all the stock of the controlled companies. Usually control could be obtained with considerably less than half the voting stock.

The decade of the 1880's saw the rapid growth of the trust form of organization. First and most famous of the large-scale business trusts was the Standard Oil Company, under the guidance of John D. Rockefeller. Some of the outstanding trusts, following the establishment of Standard Oil, were the Cottonseed Oil, the Whisky, the Linseed Oil, the Cordage, the National Lead, and the Sugar Trusts. In many cases the

trusts of this decade developed out of the less formal pooling arrangements of the preceding one. Likewise, in many cases, trusts also attempted to broaden their control through consolidations with competing companies that were not members of the trust. But the predominance of the trust form led to the establishment of lay usage of the term "trust" as almost synonymous with huge business organizations possessing a high degree of monopoly power. Thus when the Sherman Act was passed in 1890 to curb the great monopolies in restraint of trade it was generally called the Sherman Antitrust Act; and ever since "trust" has been commonly used as more or less synonymous with large-scale monopoly, whatever the technical form. We shall, however, confine the term to the "trust" form as such.

In the vast majority of cases, the great trusts represented highly centralized control by small numbers of "insiders," who apparently were primarily interested in obtaining maximum returns for themselves, at the cost of other security-holders as well as consumers. One of the most common abuses was the issuing of "watered stock." In almost all cases the securities issued to set up the controlling trusts were of a much higher face value than was justified by the assets of the trust. By selling such "watered" securities, promoters were able not only to obtain large amounts of funds to buy up control of competing companies but also to retain huge sums for themselves as commissions and salaries. In some cases, such as the Sugar and Cordage Trusts, this overissue of securities went to almost unbelievable extremes, and the promoters were the only ones to make any gains from the arrangements, while securities purchased by "outsiders" were virtually worthless. After 1890, control over the larger trusts became increasingly centralized in the hands of a relatively small group of financiers, who, through their control of the financial arrangements of the companies, retained and exercised almost complete control over their operations.

The holding companies

The Sherman Antitrust Act of 1890, adverse decisions under common law in some of the state courts, and court interpretations and popular antagonism to the "big business trust" led to a shift away from the trust form in the decades of the 'nineties and the early twentieth century. During this period the holding company became an important tool for concentrating control in business enterprise.

The holding company is similar to the trust, with one primary difference. The trust simply holds the shares of the controlled companies as trustees; it does not own them outright. The holding company buys up securities to get power over the companies that it wants to control. As the name indicates, the holding company typically owns and holds securities of other corporations; it frequently owns no land, machinery, or other such operating property itself.

To illustrate how a holding company may be utilized to obtain a high degree of control over vast amounts of capital with a minimum of invest-

ment, let us take a hypothetical example, such as is presented in Figure 26—1. Suppose there are twelve railroads, named A through L, each capitalized at \$100,000,000, of which half is bonds, one fourth nonvoting preferred stock, and one fourth common stock. Suppose a group wishes to obtain control of these railroads, but does not wish or is unable to put

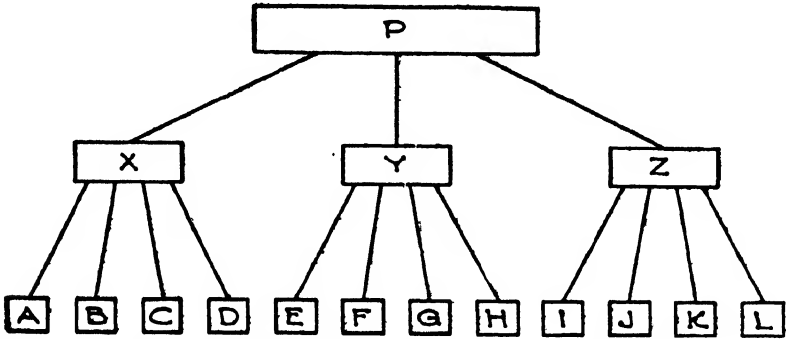


Fig. 26—1. Hypothetical holding company structure.

Total Investment of \$1,200,000,000 controlled by an investment of \$2,500,000 in the Second Degree Holding Company, P.

Each Operating Company has:	
Total capitalization of	\$100,000,000
Bonds	50,000,000
Preferred stock	25,000,000
Common stock (voting)	25,000,000

Each First Degree Holding Company has:	
Total capitalization of	\$ 50,000,000
Bonds	25,000,000
Preferred stock	12,500,000
Common stock (voting)	12,500,000

The Second Degree Holding Company has:	
Total capitalization of	\$ 20,000,000
Bonds	10,000,000
Preferred stock	5,000,000
Common stock (voting)	5,000,000

in enough money to buy control of each. The total investment required to obtain control of railroad A is just over \$12,500,000, since this is half the voting stock. A similar amount would be required for each of the other railroads, if it is assumed that there is no inertia among the other stockholders, so that half the common is actually needed to control. In this situation, the group forms holding company X to buy up control of railroads A, B, C, and D. The total capital required is \$50,000,000, since \$12,500,000 is needed to get control of each of the four roads. To raise this fund they sell, let us say, \$25,000,000 of bonds, \$12,500,000

of preferred, and \$12,500,000 of common in holding company X. Since only half the common stock outstanding is required to control company X, our group need buy only \$6,250,000 of the common themselves to hold control, obtaining the rest of the funds from the public. Then they do the same with holding company Y to buy up control of railroads E, F, G, and H, and holding company Z to buy up control of railroads I, J, K, and L. They now have control of all twelve railroads with a capital of \$1,200,000,000 through having invested \$19,000,000, the amount necessary to control holding companies X, Y, and Z, which in turn control the twelve operating railroads. Since X, Y, and Z hold the stock of operating companies directly, they would be described as "first degree" holding companies.

But our group can go still further. To control the three holding companies \$19,000,000 is necessary. Therefore they form a "second degree" holding company P, capitalized at \$20,000,000, to raise these funds. In forming company P they issue, let us say, \$10,000,000 of bonds, \$5,000,000 of preferred, and \$5,000,000 of common. Since only half the common is needed to control, only \$2,500,000 need be invested by the promoters to retain control of P, with the remainder of the securities sold to the public. With this \$2,500,000 investment, our group has control of holding company P, which does nothing but hold controlling stocks in holding companies X, Y, and Z, which in turn hold stocks controlling the twelve operating railroads. The investment of \$2,500,000 is sufficient to establish control over the entire \$1,200,000,000 set of operating railroads. If we recognize that many stockholders fail to vote their holdings, less than half the voting stock is required to maintain control in the various corporations; and the promoters would be able to accomplish the same result with an even smaller investment. And it would be possible to pyramid further with a holding company controlling company P, and so on.

This hypothetical example doubtless seems fantastic. As a matter of fact it is far from that. Numerous holding company pyramids have been built up over the course of the past two or three decades where the proportion of investment to total capital controlled is substantially smaller. For example, a recent court trial showed that one Howard Hopson, a shrewd promoter, had managed to hold control of the billion dollar Associated Gas and Electric holding company system with approximately \$100,000 of voting stock. Especially in the public utilities field, holding companies have attained an almost unbelievable complexity. Figure 26—2 shows the actual holding company structure of the great Insull financial empire shortly before it collapsed under the pressure of the depression of the early 'thirties. Although only a few other holding-company pyramids have approached or surpassed the complexity of the Insull controls, this chart indicates the possibilities of such financial arrangements, and similar though less extensive systems are still common among the public utilities today, subject to increased regulation by the Securities and Exchange Commission.

Among the familiar names of corporations established during the period

[illegible][illegible][illegible][illegible]

Answer

Indicates interest through Corp. Secor. Co. of Chi.
Inc. 111. Laventue to.
This subsidiary relations among subdividing and
operating companies.
Lower subsidiary relations.
Joint control by companies is indicated.

2) These minorities owned directly by Insull-Halsey Smart interests were supported and made fully effective by the following holdings of other Insull interests in Corporation Securities Co. of Chicago and Insull Wholly Investments, Inc.

[illegible]

Memoranda as of December 31, 1991, latest date for which data are available in Federal Trade Commission files. These memoranda list the major interstate consolidated operators of other minerals and also by the same listing family, Martin J. Imanell and Imanell Group, Inc. and their associates held key executive positions throughout the Imanell group.

from 1890 to 1904 as holding companies are the American Tobacco Company, United States Rubber Company, United States Leather Company, the International Pulp Company, American Can Company, and United States Steel Company. The period from 1904 until this country's entrance into the war in 1917 saw a marked slowing down in the formation of holding companies, along with other close-knit federations. A few huge corporations were established, for example General Motors and the Aluminum Corporation of America, but the stringent "trust-busting" campaign instituted by President Theodore Roosevelt, and the weakening of the market for new securities following 1903, together with the general low ebb of economic activity during much of that period, provided relatively unfertile ground on which to build more large business concentrations.

Business consolidations

In some cases the control of an industry in the hands of a few firms was in part accomplished through outright merging of the concerns, the previously separate firms retaining no separate identity. In some cases a growing firm absorbed smaller competitors; in other cases two or more firms were merged as one. The greater the potential importance of the economies of large-scale production and of increased controls in the marketing of products, the greater the incentives to the formation of huge combinations. In the last century most of the important large business control-units involved trusts or holding companies; more recently many companies have resorted to both vertical and horizontal combinations without these older forms.

A consolidation of business units is said to be vertical when it involves the control over two different stages in the production of one good. For example, United States Steel has bought up control of iron and coal mines supplying the needs of its steel mills. A consolidation of business units is said to be horizontal when the businesses carry on production of the same product or the same stage in the production of some product. For example, when the American Tobacco Company bought up the plants of competing cigaret manufacturers the combination was a horizontal one. The same terms apply when a company expands by setting up new plants rather than by taking over old ones. Modern industry shows a high degree of integration, both horizontally and vertically, but especially the latter. Outstanding among vertical combinations are the great steel companies, which own the mines that produce their iron ore and coal, the quarries for their limestone, the railroads and Great Lakes steamers that transport these ingredients to the mills, and the facilities for the entire process of steel making, from ingredients to finished steel parts for use in other industries. Perhaps the purest examples of horizontal integration have been in the field of investment banking, where large numbers of firms have gradually combined to make a few financial giants that for decades have dominated the floating of new corporate issues.

The late 'twenties of this century saw a boom in the development of business consolidations comparable to the "wild" era of trusts at the end of the last century. Like the earlier orgy it was largely the work of speculative promoters who saw profits for themselves in their manipulations.

Loose-Knit Federations Privately Organized

Loose-knit federations are arrangements for cooperation between separate firms seeking to follow a common policy and to avoid active interfirm competition. They differ from close-knit federations in that each firm remains officially under separate ownership and control. The framework of collaborative action has taken many different forms; but loose-knit federations of business firms aiming at monopolistic controls over the markets are probably as old as the exchange economy, which is very old indeed.

Informal agreements

The great expansion of the domestic market following the Civil War, coupled with rapid technological advances conducive to low-cost, large-scale production, laid the basis for the ensuing growth of large enterprises and concentration of control of an industry in a few firms. The general downward trend of the price level through this period, following the inflated prices of the war, put a strong pressure on businessmen to find ways of maintaining profits and of resisting the downward pressure on prices. The first result of these conditions was very strenuous price competition in many industries. Out of such bitter competition came informal price agreements, and the more formal "pools" to be discussed in the next section. The informal price agreements were in some cases secret, in others open; they were sometimes "gentlemen's agreements," which were oral only; in other instances the agreements were set down in written form, occasionally with some provision for enforcement by penalties within the group or with threats of competitive retaliation for non-conformance. Since these were for the most part regarded as agreements "in restraint of trade" they were not enforceable at law. They lasted only as long as the participants found it profitable to comply, and no longer. The following are examples of industries in which simple price agreements were important during the period immediately after the Civil War: anthracite producers, bridge builders, gunpowder manufacturers, meat packers, railroad companies, tile manufacturers, and, in local markets, druggists, coal, ice, lumber, and tile dealers, and truck farmers.¹

The development of informal agreements in the twentieth century has involved much more subtle and complex procedures. There has been a development of "conventional practices" in both pricing and market sharing. Some brief examples will illustrate this situation.

¹ Haney, Lewis H., *Business Organization and Combination*, The Macmillan Co., New York, 1913, Chapter 10.

Conventional sharing of the market for packing-house products was established as early as the turn of the century, and has continued to the present. Five, and more recently, four, large packing companies have continuously dominated the market in the buying of livestock, and although there have been some shifts among them the shares received by each have remained amazingly stable. By agreements as to what proportion of the total purchases of livestock should be made by each company, they automatically provided a technique for sharing the market in selling the processed meats as well. Price competition among the "Big Five" has generally been confined to relatively nonaggressive practices.

The same sort of conventional sharing on a live-and-let-live basis has been and still is fairly common among the major oil companies. Testimony before the Temporary National Economic Committee in 1939 revealed that established conventional shares existed in many areas, even though the shares had never been agreed upon and although no company felt any responsibility for maintaining any share except its own. Typical were the statements of oil men that they were interested in seeing that there was no "poaching" on their preserves, but that disastrous price wars were strongly to be avoided.

Especially strong today is conventional market sharing in the construction industry, as the evidence before the Temporary National Economic Committee reveals. Among the big builders of industrial plant and government buildings this is accomplished in many areas through collusive bidding on contracts. By agreeing before each bidding who is to get the contract, the successful bidder can place his bid at a highly profitable level without fear that competitors will underbid him, since by prearrangement they will bid even higher. On the following contract another contractor will have his turn, and so on. As with all other price-raising, market-sharing schemes, collusive bidding can work successfully only so long as all participants can agree on an acceptable means of market sharing. As long as the number of contracts received by each provides reasonably profitable operation, the arrangement is likely to continue in spite of probable overinvestment in the industry. But once the shares fall drastically because of diminished demand or entrance of new firms, open price competition is likely to replace conventional market sharing.

Informal price agreements, based primarily on tacit convention, have been increasingly associated in recent years with competing for customers through aggressive advertising. One of the most extreme examples is the cigaret industry. Studies of the industry reveal that relative shares of the different major companies in the total sales of cigarets have fluctuated quite consistently in response to shifting relative expenditures on advertising.² It is extremely difficult and very risky for new firms to make

² U. S. Securities and Exchange Commission, *Survey of American Listed Corporations*, 1940, Vol. 1, p. 16.

any major sallies into this industry. Whenever one company changes the price charged, the others follow immediately, usually the same day.³ Commenting on this situation the Federal Trade Commission stated that "although these prices may not be established collusively, there seems to be an unwritten rule that any price change will be followed."⁴ However, these prices have been highly stable. Manufacturers' prices of cigarets changed only eleven times in the twenty-year period from 1919 to 1939; they did not change at all from October 1922 to April 1928, nor from January 1934 to January 1939.⁵ Changes in cigaret prices have shown little relation to either prices of leaf tobacco or general commodity price levels.

The tobacco industry also illustrates informal agreements in buying. Advertising is not important in this sphere, however, and hence practices have taken other forms. The Federal Trade Commission reports that the major companies have depressed tobacco prices by making all their purchases through one buyer, by buying only a certain per cent of the offerings, by timing purchases in such a way as to force sellers into a "distress" position when they most want to make sales. Some of these practices have been fostered through the Tobacco Association of America, a trade association of firms in the industry.

Price and market-sharing agreements on informal bases are most common and most successful in industries in which there are only a few firms or in which one firm is clearly dominant and takes the initiative as leader, others tacitly following its policies. The agreements are frequently facilitated, as in the case of tobacco buying, by trade associations; they are also commonly supported by interlocking directorates. Both of these will be discussed in subsequent sections.

The "pools" of the nineteenth century

"Pools" have been distinguished from simple agreements in that "pools" provided some kind of administrative agency for the execution and enforcement of the agreement. Such agencies exist today, but commonly the word "pool" has been used with a historical connotation to refer to the earlier cases. The informal agreements following the Civil War rarely included the market-sharing features common to such agreements in more recent times; agreements for the sharing of markets were primarily carried out through the more formal machinery of a pool.

There were many variations in the structure and programs of these early pools. For example, the Michigan Salt Pool provided a formal selling agency. All members turned their outputs into the pool, whose managers then sold the salt at the most advantageous prices, after which the proceeds were distributed among the pool members. Such agencies

³ Federal Trade Commission, *Report on the Agricultural Income Inquiry*, 1938, Part I, p. 447.

⁴ *Ibid.*, p. 464.

⁵ Malott, D. W., and Martin, B. F., *The Agricultural Industries*, McGraw-Hill, New York & London, 1939.

were also maintained by the producers of blue stoneware, coal, lumber, Manila and fiber paper, petroleum products, shade-rollers, wallpaper, window glass and wooden dishes. In other cases, especially in agreements among competing railroads, the earnings of the members were pooled and a definite proportion of the total pool was allocated to each road, thereby removing the incentive to price cutting and non-price competition among the roads. Still another scheme was exemplified by the Standard Envelope Company, a pooling arrangement wherein each member had to pay into the pool a tax on all output above a certain amount in order to restrict output that might have the effect of depressing price. And in some cases the pools divided the total market into geographic segments allocated to the various pool members; most famous among such cases was the Addyston Pipe and Steel Company and five other corporations in a pooling agreement that was both a selling agency and a device for dividing the market for cast iron pipe. The Steel Rail Pool formed in 1887 was one of the most famous of those establishing direct output quotas for members.⁶

Trade associations as the bases of modern "pools" or "cartels"

To a much greater extent than is commonly imagined, recent loose-knit federations resemble the pools just after the Civil War, though the word "cartel" is more commonly used to describe them. In many foreign countries, especially Germany, these cartels have been openly fostered by the State; but in the United States this has not generally been the case. They have functioned less openly here, since they have generally been regarded as illegal, though there are important exceptions. Trade associations, which are perfectly legal in themselves, have frequently been used as agencies for the administration of policies of collaboration that in fact amounted to pooling arrangements.

Almost every modern industry of any importance has its "trade association" or "institute." The activities of these groups, formed by producers in the industry, have covered a wide range from time to time and from industry to industry. The trade association was developed to aid the producers in the industry. Toward this end it commonly carries on advertising activity and other public relations work, maintains commercial research, conducts investigations into legislation and other happenings affecting the industry, attempts to influence legislation in behalf of the industry, provides information for its members on methods of selling, accounting techniques and other such matters, and any other helpful activity. But the trade association sometimes has another purpose, that of encouraging "cooperation" among producers with respect to price and output policies.

⁶ Pooling arrangements were also common in the sale of bathtubs, bottles, brass, cordage, cotton bagging, cotton thread and yarn, explosives, iron and steel, meat, nails, naval stores, sugar, tobacco, whisky, and many other products. (See Haney, *op. cit.*, Chapters 11 and 12 and Jones, Eliot, *The Trust Problem in the United States*, Macmillan, New York, 1922, Chapter 2.)

In the course of its many activities to aid its individual members, it is almost inevitable that the trade association will develop some degree of power to influence and coordinate the policies of firms in the industry. Especially is this true when the association collects and distributes among the members detailed statistics on the costs, output, and prices of each member. Given such a set of data, it is a short step for members to follow cooperative price policies intended to hold price above some standard cost figure. To this end standardized systems of cost accounting have often been suggested to member firms. Once price maintenance is established, the association provides the logical channel for developing and administering market-sharing plans.

That the trade association has in many cases led to a more efficient and economical ordering of production cannot be doubted. On the other hand, that one of the primary purposes of a very large number of trade associations has been the avoidance of price cutting by members is equally beyond doubt. The trade association provides an instrument for price maintenance and share-the-market arrangements, and it has been so used in a large number of cases. Trade association officials and members are fond of saying that they want to retain the competitive system but eliminate its "wastes and disastrous price wars." Designed partly as a compromise between "disastrous competition" and detailed government control of industry, the trade association has served well to establish such a "compromise" situation throughout the modern American economy. For its part, the state first ignored trade associations, then encouraged them, then restrained them, then adopted them as its chosen instrument of industry control in the National Industrial Recovery Act of 1933, and then proceeded to challenge them once more.

Cooperative selling agencies

Market sharing by pooling the total output of the participating producers for sale with subsequent distribution of the proceeds was common in the nineteenth century. By its control of sales, the central selling agency was able to maintain "profitable" prices. Perhaps the most prominent among such arrangements recently has been that of the producers in the bituminous coal industry, set up in 1932 to market the entire output of participating producers. The central agency carried on sales campaigns and allocated orders among the participating producers. Likewise, in agricultural industries cooperative marketing has come to occupy a position of prominence, especially in the fields of dairy products and citrus fruits. Marketing cooperatives of dairy farmers in areas surrounding major cities were in many cases developed partly to increase the bargaining power of farmers dealing with big and powerful milk distributors; but these cooperatives have in some cases joined with distributors in elaborate schemes of price control and output restriction. In some of the cooperatives express restriction of the production of each farmer is attempted; in others, farmers apparently merely avoid obviously overproducing the market demand at the agreed price. The marketing

cooperative sells milk in the most advantageous manner and divides the proceeds among the participating farmers in accordance with a previously determined plan. The way in which citrus fruit producers have used such cooperative organizations to create bottlenecks in the marketing of fruits will be described in a later section.

Interlocking directorates

When certain men or banking houses are represented on the board of directors of several different corporations, it is relatively easy to control the policies of these corporations in concerted action to raise or maintain selling prices, to lower buying prices, and to employ weapons to prevent the entry of new firms into particular industries. This duplication in directorships is described as "interlocking directorates."

Some idea of the interrelations among corporations through interlocking directorates at the present time can be obtained by examining the 200 largest nonfinancial and 50 largest financial corporations in this country. In 1935 only 35 of these corporations had no director in common with other corporations on the list. Between them, 400 men held nearly a third of the directorships in all these corporations; 1,000 men held over half of them. One corporation, the Western Union Telegraph Company, interlocked with 35 other corporations on the list. Of the 250 corporations, 151 companies, whose assets amounted to almost three fourths of the total, were interlocked with at least three others of the group.

That interlocking directorships provided an inviting avenue for the formulation of corporate policies in restraint of trade is obvious; yet it is by no means certain how important this characteristic of the modern corporate community has actually been. If the interlocking directors are relatively inactive, their presence on several boards may mean little; if they play an active role in the control of competing corporations the likelihood of policies to restrain competition is very great.

Publicly Fostered Cartelization

Codes of "fair competition" and the National Industrial Recovery Act

Passed in the spring of 1933, the National Industrial Recovery Act, usually called the NIRA, had as its primary purpose the stimulation of recovery from the unprecedented depression in which the nation found itself. One of the major provisions of this act was the requirement that every industry should formulate a code of "fair competition" that, when approved by the President of the United States, would be binding on all members of the industry. But the phrase "fair competition" was on the whole a very misleading one; it in many cases was in effect just an attractive way of describing a policy of eliminating a large share of competition (especially price competition).

The code provisions of the act were a definitely formulated invitation to trade associations or other industry groups to place before the President

as far-reaching plans as they could agree on for the centralized control of their respective industries. The actual codes approved represented the demands of the trade groups somewhat modified by the National Recovery Administration, though heavy reliance on the trade associations themselves as administrators of the codes gave the associations a further opportunity to influence policies under the codes. Usually the "big fellows" determined the policies, and small businessmen were often the "goats." Almost all codes provided for some form of minimum price setting and most of them for some direct or indirect control of output. In many cases stated minimum prices were provided; in virtually all, selling below some standard "cost of production" was prohibited. In many cases sellers were required to adhere to "open price" policies, that is, to publish changes in their prices in advance of making these changes, so as to provide full information to competitors. Some codes provided for full control of output, but the administration refused to permit this provision to be used except in the cases of lumber, petroleum, and copper, although in virtually all cases lesser restrictions on production were approved, one of the commonest being a maximum number of hours per week that plants might be operated. The restriction on working hours for laborers also tended to serve the same purpose.

NIRA was an amazing study in internal contradiction. It gave to industry groups the power to establish cartels controlling both price and production, protected by law; and it then set up a consumers' division to see that these monopolies did not raise prices "excessively," and administratively admonished producers not to violate the laws against practices in restraint of trade. As a development in the concentration of control and power in the American economy, it occupies a place of major importance. While it lasted, NRA organized virtually all of American industry into a series of compulsory cartels. Adherence to "fair" prices was not suggested; it was required. Output restriction was often in the same category. Other types of "cooperation," such as open price systems, pooling of statistical information on costs, sales, and output, elimination of much non-price competition, and even control of new investment in the industry were common in the codes. NIRA was unanimously declared unconstitutional by the United States Supreme Court in 1935, but the influence of the Act lingered long after the famous Blue Eagle of NRA had vanished from sight. A number of trade associations sought to prevent an exodus from the promised land of "cooperation" by declaring in favor of continued adherence to the codes no longer legally enforced. But far more important was the less open carryover of adherence to price and output policies established formally by the codes.

Recent cartelization under federal auspices

Analogous to the privately controlled schemes of big business is the most gigantic formal market-sharing arrangement of all time, the production and price control mechanism of the governmentally supported Agricultural Adjustment Administration. Under this arrangement, which

is discussed in detail in a later chapter,⁷ each participating farmer was given a crop quota based largely on the previous years' outputs from his farm, and price was supported by the federal government. Penalties in the form of withdrawal of bonus payments and ineligibility for crop loans were imposed on farmers exceeding their quotas. Although governmentally sponsored, the AAA was simply a huge cartel arrangement, leading to essentially the same results as to production and price policies as other cartel arrangements. In its essential features this program still remains. A purely competitive industry is thus converted, through governmental intervention, into something resembling the pools formed by business groups in industries of only a few firms.

Centralized control over the marketing of agricultural products was also legalized and fostered by the Agricultural Marketing Agreement Act of 1937. Here the Department of Agriculture comes directly into conflict with the Antitrust Division of the Department of Justice, to be discussed in Chapter 28. The most striking uses of this power have occurred in the marketing of citrus fruits and of milk.

Even greater is the support given to coordinated fixing of high prices on bituminous coal by the Bituminous Coal Act of 1937; here government steps in not merely to assist in supporting high prices, as under the AAA, but actually to determine and enforce high minimum prices. The Coal Act is a direct outgrowth of the NIRA codes, and the Agricultural Marketing Agreement Act is its second cousin.

Controlling and Creating Bottlenecks

Thus far we have focused attention on the organizational framework through which power may be built. But how do these organizations themselves attain and maintain power? There are usually certain essential points in the functioning of an industry that are more easily dominated than others; by controlling these points, which then become bottlenecks, the leading producers of oil or sugar or tobacco or meat or some other product may strengthen and consolidate their positions of dominance. Where these bottlenecks may appear will differ from industry to industry. Only a sample of some of the most important can be examined here: (1) transportation facilities, (2) marketing channels, in both the buying of productive services and the selling of products, (3) raw material and power sources, (4) patents as bases of control over the exploitation of technological improvements, (5) labor supply.

These controls are obtained in part simply by purchase, but commonly they are attained and maintained by unfair competitive tactics. Although there are many honest and fair businessmen, unfair competitive tactics are nevertheless very common and they take many forms. They include discriminatory price cutting to eliminate rivals, use of strong bargaining positions to secure support of banks or railroads in discriminatory actions, use of spies and sabotage of competitors, gross misrepre-

⁷ See Chapter 57.

sentation of the products and positions of rivals, threats of withdrawal of patronage if a firm entering into negotiations with the monopolist or the oligopoly group should have any dealings elsewhere, and so on. Some of these practices, especially manipulation of railroad rates, were especially rampant in the era of the growth of the big trusts, but many of them retain great importance today.

Controlling transportation facilities

Transportation facilities may be controlled by a dominant group in a particular industry either through ownership of these facilities or through various types of pressures that may be brought to bear on those who do own and operate them. Such controls are commonly associated very closely with unfair methods of competition.

John D. Rockefeller is the most famous and was the most successful of those manipulating railroad rates in the late nineteenth century. The original advantages of the Rockefeller interests were obtained primarily through playing off against each other the three major railroads connecting the Pennsylvania oil fields with the East. Since transportation costs were more important than refining costs, competitive rebates obtained by Standard from the Pennsylvania, the Erie, and the New York Central provided a major cost advantage over competitors not obtaining such rebates. Having obtained substantial control over the oil markets during the early 1880's, Standard was in a position to exercise even stronger pressure, both on competitors and on the railroads. At one time Standard held an agreement whereby it received rebates on the shipments made by *other* oil producers, as well as on its own shipments, though this provision was quickly declared void by the courts. The development of pipe-line transportation introduced a new element into the picture; Standard, having established itself in a virtually impregnable position through its railroad transportation arrangements, lost no time in obtaining almost complete control of the country's pipe lines. In the pipe-line struggle as in the earlier rail struggle, transportation rates were slashed to virtually zero, and even below in some cases, and oil and kerosene were almost given away by the rival companies in the fight to obtain the market.

In more recent years the major petroleum refining companies have maintained a controlling position in the industry largely through their ownerships and operation of the pipe lines. This development has been particularly important in the petroleum refining industry because transportation is such an important part of the cost of petroleum production and because of the development of a specialized means of transporting the product that is much cheaper than alternative methods. Although the pipe lines have been under the jurisdiction of the Interstate Commerce Commission since 1906, little has been done until very recently to control their rates and to regulate other practices. Independent producers have been unable to ship by pipe line because of some of the policies of the major companies controlling the lines. For example, "minimum tender requirements" are requirements stating the minimum number of barrels which

the pipe line will accept for shipment, and these requirements on the main trunk lines carrying crude oil have frequently been as high as 100,000 barrels; but the oil companies owning the pipe lines do not apply this minimum to shipments for their own accounts. Rarely can an independent producer arrange to ship in such large quantity. Moreover, the problem of terminal storage facilities constitutes an obstacle. Without such facilities near the source of the oil the independent producer is unable either to hold off in order to get a more favorable price for oil sold at the wells or to accumulate enough oil to meet legal tender requirements in shipping it elsewhere. In order to ship his oil (if he does not sell it at the well), the independent producer must also have storage facilities at destinations since pipe lines usually require "that crude oil will not be received for shipment unless provision is made for immediate acceptance in the consignees' tanks at destinations."⁸ Storage facilities are not available to all who want to use and pay for them; also they are owned by the major companies. Most independent oil producers could not afford to provide storage facilities of their own, since they would not use their facilities enough. Control of oil pipe lines and storage facilities on these lines thus gives the major companies a firm clutch on the entire oil-refining industry, which could be remedied only by a reorganization of the industry or by strict regulation by the Interstate Commerce Commission. Some action to remedy the situation has recently been undertaken.

Although the most notorious examples of controlling and creating transportation bottlenecks have occurred in the petroleum industry, because of some of the special conditions of that industry, there have also been important instances in other industries. Getting rapid and efficient service in the transportation of perishable goods is particularly important.⁹ The control of the banana industry by just two major concerns at the present time is an interesting illustration. Nine tenths of the bananas produced in quantity for export come from Central and South America and the West Indies. Most of this trade is handled by the United Fruit Company and the Standard Fruit and Steamship Company; the former

⁸ *Report on Pipe Lines*, Part I, House Report No. 2192, 72nd Congress, 2nd Session, p. LXVIII.

⁹ For this reason the ownership of refrigerator cars by the Big Five in the meat-packing industry was at one time an important tool of control. In 1917 these companies owned 91 per cent of the refrigerator cars properly equipped for shipment of fresh meat. Icing stations between St. Louis and Chicago and the Atlantic seaboard were owned by three of the Big Five. All cars, including those owned by independent packers, were iced at these stations. The big packers thus gained and made use of information concerning competitors' business. In addition, the cars owned by the big packers received preferential treatment by the railroads; they were carefully handled, promptly returned, and used only for shipment of the packers' own commodities. Cars owned by smaller packers were subjected to extreme delays in return, e.g., six months to nine months for a trip from St. Louis to New York and back. The railroads used independents' cars to ship other commodities, for example onions. FTC, *Report on the Meat Packing Industry*, submitted June 24, 1919, Part I, pages 40-41.

was responsible for 60 per cent and the latter for 30 per cent of the bananas coming into the United States in 1936. The United Fruit Company

, . . . owns or leases 3,500,000 acres of land. It operates a fleet of a hundred ships. It runs all but one of the banana-carrying railroad lines in Central America. It owns the Tropical Radio Telegraph Co., which offers the only telegraph service between the United States and Honduras. It operates docks and stores, hospitals and hotels, . . . The United is in a position to exact high railway rates of other shippers and to prevent other ships from loading bananas by giving preference to its own ships at its docks. It is said to have chartered cargo space, which it did not use, on the boats of other lines for the purpose of preventing other shippers from reaching the market.¹⁰

Controlling marketing channels

Controls over marketing facilities have been common among oligopoly and oligopsony groups in many industries. This is an especially important technique for raising and maintaining prices on some agricultural products.

One of the most elaborate systems was that developed by the meat packers prior to World War I. Most of the market stockyards were originally established and managed by railroads and by other interests outside the packing industry. About 1890 some of the leading stockyards offered inducements to packers, such as bonuses of land, buildings, stockyard bonds, stock, and sometimes cash payments, to establish plants in their yards. Thus the Big Five soon obtained a substantial interest in the most important stockyards. Despite the huge bonuses granted to them, the Big Five, as they grew and became more powerful, began to demand a still larger share in the ownership and management of the yards. Stockyards without packing plants were a poor investment; hence, concessions were granted.¹¹ In gaining control over these stockyards, the big packers used threats to move business elsewhere to force dividends to themselves. Sometimes complete reorganization of the yards was demanded to give packers a controlling interest. Because of the dominance of big packers new stockyards were not opened up without their approval, and therefore their dominance.¹² "By excluding banks and cattle-loan companies, except those controlled by the packers, from convenient locations near the yards, the Big Five controlled livestock credit in the principal markets, which control in turn gave them power to force livestock on the markets through calling outstanding loans. In addition, control of the yards enabled the packers to establish undue

¹⁰ TNEC Monograph No. 21, pp. 101-102.

¹¹ By 1917 the Big Five either severally or jointly owned the capital stock of public stockyards as follows: 100 per cent—Milwaukee, Denver, Dallas, Jacksonville; 90 per cent—New York, Pittsburgh; 80 per cent—Jersey City, Oklahoma City, Portland (Ore.), St. Joseph, East St. Louis; 70 per cent—Sioux City; 60 per cent—St. Paul, Ft. Worth, El Paso; 50 per cent—Kansas City; 23-40 per cent—Philadelphia, Louisville, Omaha, Wichita. Chicago figures are not available, but Armour alone owned 19 per cent. FTC, *Report on Agricultural Income Inquiry*, 1938, p. 202.)

¹² FTC. *Report on the Meat Packing Industry*, Part I, pp. 132-133.

control over the commission houses through which most of the livestock entering the market was sold. The commission houses were tenants of the yards, and therefore were dependent on the packers controlling the yards."¹³ For instance, before commission houses were permitted to sell livestock in packer-controlled yards, they had to sign an agreement to turn over all dead animals to stockyard management at prices dictated by the management.¹⁴

By gaining control of the stockyards, the big meat-packing concerns were in a very strong position to initiate and enforce restrictions of production in order to push up and to maintain at high levels the prices at which they sold their products, and to depress the prices paid for animals purchased.¹⁵

Cooperative selling agencies in some industries have made possible centralization of control in the marketing of goods produced by thousands of different firms. Such organizations may become highly monopolistic in their controls and their policies. An interesting example is provided in the citrus fruit industry since 1925. The California Fruit Growers' Exchange, popularly known as "Sunkist," is the largest of the marketing cooperatives in the citrus fruit industry; 14,500 out of the 18,000 citrus fruit producers in California and Arizona are members of the organization. In attempting to maintain high prices for citrus fruits the organization runs into the difficulty that these high prices encourage expanding output; but to control the production of the thousands of fruit farmers would be an almost impossible task. It has therefore restricted the flow of fruit through the markets by a number of techniques, including the allocating of shipments among the members, surplus fruits being diverted into plants producing by-products. Its vulnerable spot is its inability to force membership in the exchange on the fruit producers, but despite this

¹³ FTC, *Report on Agricultural Income Inquiry*, p. 203.

¹⁴ *Ibid.*, p. 207. In addition to controlling the stockyards, the big packers controlled a large proportion of branch wholesale houses and cold storage plants, and had large interests in public cold storage warehouses. This domination was almost as great in 1935 as in 1917 when an intensive investigation of the packers was made by the Federal Trade Commission. (*Ibid.*, p. 137.)

¹⁵ Another interesting example of control over marketing channels is provided by the motion picture industry. This situation was described briefly in an article entitled "Trade Barriers Created by Business," written by Corwin D. Edwards and published in the *Indiana Law Journal*. Mr. Edwards writes as follows: "Major producers of moving pictures . . . control affiliated chains of moving picture theatres through which they may be sure of a wide distribution of their product. Until the recent consent decree in the case of *United States vs. Paramount Pictures, Inc., et al.*, it was customary to offer a fifty-two weeks' supply of films in a single contract, so that independent theatres wishing to deal with the major companies were required to contract for a block of pictures so large as to leave them without demand for films from other studios. The majors gave preference to their theatres and to certain powerful independent chains by permitting them to show new pictures first and sometimes by supplying them with better films and charging them lower rentals. The independent theatre was at a disadvantage in competing with the controlled theatre and with the independent theatre chain. The independent producer was likewise at a disadvantage in marketing his product." (December, 1940, p. 160.)

weakness Sunkist and the other large producers' cooperatives in the industry were found to "have and exercise, directly and indirectly, the power to influence and dominate the marketing policies and practices of the Auction Companies, the prices at which and the channels through which receivers, buyers, and consumers throughout the United States obtain citrus and deciduous fruits distributed and sold in interstate trade and commerce."¹⁶

Controlling raw material and power sources

In a monograph on "Competition and Monopoly in American Industry," published by the Temporary National Economic Committee in 1940, there is a listing and brief discussion of those industries that approximate "complete monopoly" in the United States. Most of these are cases of exclusive control of basic raw materials and/or of basic patents; some are public utilities or quasi-utilities (to be discussed in a later chapter). Some quotations from comments in this publication illustrate this situation with regard to raw material controls very clearly:

For more than 50 years, the Aluminum Co. of America has produced 100 per cent of the Nation's output of alumina and virgin aluminum ingot. For some 30 years it has been reported to own or hold more than 90 per cent of the commercially available supply of the raw material, bauxite. It has used 100 per cent of the bauxite produced in the United States.¹⁷

The International Nickel Co. of Canada, Ltd., owns more than nine-tenths of the world's known reserves of nickel. The company produced more than 92 per cent of the world's output of nickel in 1929.¹⁸

Molybdenum is an element which finds its principal employment either in competition or in combination with other alloying metals, in the production of steels of exceptional toughness and strength. In Bartlett Mountain in Colorado, the Climax Molybdenum Co. owns 95 per cent of the world's known store of commercially workable deposits of this metal.¹⁹

The best known of these examples is the Aluminum Company. Its controls are based not only on possession of bauxite but on other foundations as well. The control in the American market has been maintained in part by tariffs that keep out aluminum produced in other countries from bauxite deposits outside the United States. The manufacture of aluminum requires very great power facilities, and the Aluminum Company has succeeded in obtaining and maintaining controls of these facilities wherever there has been any competitive threat (until the recent beginnings in the Tennessee Valley under Government auspices). Even during World War I the Aluminum Company maintained its monopoly position. It was so powerful that it could force power companies that it did not own to refrain from selling electricity to any potential competitors; and banks were afraid to finance any new enterprise.

¹⁶ Criminal Indictment No. 15167 in the District Court of the United States for the Southern District of California, Central Division, December 17, 1941, p. 16.

¹⁷ TNEC Monograph No. 21, p. 69.

¹⁸ *Ibid.*, p. 79.

¹⁹ *Ibid.*, p. 81.

Patents

One of the most important of the foundations of monopoly and oligopoly power in American industry has been patents. Examples are many. Such patent controls in the early history of the aluminum industry were what made possible the establishment of the Aluminum Company of America in a position of continued control in the industry; during the twenty years' duration of the patent on the electrolytic process the company acquired ownership of the bauxite deposits and the technique, personnel, and financial resources needed to maintain its exclusive position. Other outstanding examples of monopolistic controls based on patents are the United Shoe Machinery Company, the control by the Hartford-Empire Company over the glass-container machinery, the concentrated control of the production of electric lamps, the production of electric counting machines, and the production of radios.

The simplest aspect of monopoly control based on patents is the outright refusal to license other producers to use a basic patent held by one concern or one small group of concerns. In some cases a few firms holding different basic patents get together and pool their patents, excluding all outsiders from the use of any of the pooled patents; major concerns in radio production created such a pool, which action made it necessary for any outside firm entering the industry to develop substitutes for all aspects of radio production at the same time before it could get under way.²⁰ The buying up of patents on substitute processes is also a common practice, notably exemplified by the glass-container industry.

The case of the United Shoe Machinery Corporation (and the two major companies producing electric counting machines) illustrates vividly (1) how patents may provide a continuing basis of control if the firm first in the ascendance continues to develop a series of new improvements that keep its patent controls always one jump ahead of what anyone else could obtain and (2) how patents on one process may be used to extend control into other spheres of production as well. The company was founded in 1899, as a merger of seven concerns that owned patents covering all of the important shoe machinery manufactured in the United States. It quickly acquired the fifty remaining patents and by 1911 it was manufacturing between 95 and 100 per cent of the shoe machinery made in this country. The company leased its machines to shoe manufacturers; it did not sell them. By the use of "tying leases" it required that companies using any of its machines should not use any machines made by any other producers, nor could its machines be used on shoes that were processed at any stage by machines from some other firm. This enabled the company to extend its exclusive control into areas of production where it had previously been in active competition

²⁰ *United States vs. Radio Corporation of America*, Consent Decree, Nov. 21, 1932.

with other manufacturers. This situation is well described in the TNEC monograph to which we have already referred: ²¹

The shoe manufacturer, who could obtain a lasting machine only by leasing it from the United Shoe Machinery Corporation, was compelled to turn to it also for his welter, stitcher, and metallic fastener, and the independent producers of those machines were robbed of their customers. The device operated also to continue far beyond the statutory 17 years the protection afforded the company by its patents. As long as any one of these patents granted it the exclusive right to produce a single machine, the tying clause in its contracts extended its monopoly to each of the others.

Although the company has been attacked in the courts and ordered to "cease and desist" from the use of tying leases, it remains in an almost completely exclusive position as manufacturer of shoe machinery, market for patents, and provider of servicing on these machines.

Another interesting example of patent controls is afforded by the recent manipulation of the patent for "tetra-ethyl," which when added to "straight" gasoline makes that gasoline comparatively "knockless." The Ethyl Gasoline Corporation is owned jointly by the Standard Oil Company of New Jersey and the General Motors Corporation. In order to use tetra-ethyl, refiners must obtain a license from the Ethyl Gasoline Corporation, and such licenses have been issued to all but one of the 124 leading refiners of gasoline. Until it was attacked by the Anti-trust Division of the Department of Justice in 1938, the license was issued under a contract that the licensee would sell his "antiknock" gasoline at a fixed differential over "regular" gasoline, and that he would sell it only to licensed jobbers. The jobbers were therefore effectively forced to obtain licenses from the Ethyl Gasoline Corporation, and these licenses were issued under contracts requiring that the jobber comply with a code of "fair competition," which in effect bound him to follow the prices set by the major oil companies. This eliminated competition in the sale of gasoline by the jobbers.²²

Controlling the labor supply

Restraints of trade by labor groups are extremely important and will be examined in detail in Chapter 32. For the present, however, we are concerned only with how business groups may be able to manipulate labor situations (frequently in collusion with labor unions) in such a way as to establish and maintain positions of dominance, excluding outsiders.

Restraints of this kind are especially common in agreements between

²¹ TNEC Monograph No. 21, p. 73.

²² "In March 1940 the United States Supreme Court found the defendant guilty of violating the anti-trust laws, asserting that the corporation, by the leverage of its licensing contracts resting upon the fulcrum of its patents had built up a combination capable of use and actually used as a means of controlling jobbers' prices and the suppression of competition among them. Since the regulation of prices and the suppression of competition among the purchasers of the patented articles was not within the limits of the patent monopoly the practices of requiring jobbers to take out licenses was enjoined." (*Ibid.*, p. 161.)

materials dealers and building contractors and trade-unions, by which the workers undertake to boycott outside dealers and contractors in return for analogous favors rendered them. The agreement of the anthracite producers and the United Mine Workers is another important case.²³

One more example may serve to sharpen up this point. In 1941 the Antitrust Division of the Department of Justice brought an indictment charging restraints of trade in milk distribution in a number of cities, including Dubuque, Iowa. One of the local dairies in Dubuque had attempted to lower distribution costs by offering a substantial reduction of price on deliveries of more than one bottle of milk. Thereupon the other creameries got together with the teamsters union to prevent the delivery of the low-priced milk. They accomplished this primarily by strikes of workers, and by fining union members delivering such milk.²⁴

Discriminatory Price Cutting

Discriminatory price cutting means the cutting of prices in selected spots and to selected groups of buyers. It may be undertaken not merely to increase trade, but to crush competitors. Discriminatory price cutting is only one of many methods of unfair competition, a number of which have already been mentioned in connection with previous discussions. It is sufficiently important, however, to deserve special attention. Examples are almost unlimited, in both historical and current situations.

Some examples in the history of the oil-refining industry

One of the most striking examples is afforded by the petroleum-refining industry. At the end of the nineteenth century petroleum refining was almost completely controlled by the Standard Oil Company. Having obtained a commanding position, Mr. Rockefeller and his associates left no stone unturned to support and widen its control. Almost all kinds of unfair tactics are illustrated in the history of this company. With its vast resources it could afford to undertake price wars wherever competition appeared, slashing prices to figures impossible to meet without taking terrific losses. At the same time Standard could hold its prices at monopolistic levels in other areas, thereby maintaining its profits and thereby its ability to fight price wars where necessary. Strong coercion appears to have been used to force rivals to give way in case of conflict, and Standard gradually absorbed more and more of the industry's refining as well as transportation facilities. At the peak of its control, in the early 1900's, it is generally held that Standard controlled between 85 per cent and 90 per cent of both the country's refining and pipe line facilities. As a result of this control, Standard consistently showed enormous profits, both from the sale of oil products and from its trans-

²³ See footnote 5, Chapter 27.

²⁴ Edwards, Corwin D., "Food Processors and the Anti-trust Laws." A speech before the Associated Grocery Manufacturers of America, Nov. 7, 1941, p. 7. (Mimeographed.)

portation facilities. Few trusts were able to boast success even vaguely approaching that of the Rockefeller interests.

Hearings before the Temporary National Economic Committee in 1940 reveal similar practices in the oil industry today, though that industry is now controlled not by one firm alone but by a cohesive oligopoly group owning the oil pipe lines. Discriminatory price cutting has been rampant, and pressure on retail dealers to confine their purchases to the major companies has resulted in the elimination of many small rivals.

Some examples in the meat-packing industry

Discriminatory price cutting in the meat-packing industry has been carried out systematically by a collaborative plan executed by the Big Five, especially in the first two decades of the twentieth century. The Big Five, or as many as happened to be represented in a given territory, arranged to cut prices in rotation, a day or a week at a time, to crush independent operators. Thus the burden, distributed among the packers, was light, but competition was effectively crushed.²⁵

Along with discriminatory price cutting big firms in the meat-packing industry have attempted to drive out some kinds of competition by discriminatory bidding up of the prices paid for some of the products they buy. This procedure was mentioned in Chapter 10 in connection with the attempts of the packers to crush the business of cooperative creameries in order to retain a more powerful position in the buying of creamery products; the big companies offered much higher prices for creamery products in areas where sellers were independent, thus checking the growth of the cooperatives.

The program of the packers has been fortified by the extreme diversification of products in the meat-packing industry. The big companies controlled not only a large share of the processing of meats and meat by-products, but also of some of the substitutes for meat products. The creamery situation just cited is an example. They also sank their roots deep in the distribution of wholesale groceries.²⁶

²⁵ FTC, *Report on the Meat Packing Industry*, 1919, Part I, p. 68.

²⁶ This situation is shown clearly in the following data:

1. Control in meats

In 1888, Swift & Co., Armour & Co., Hammond & Co. (later acquired by Armour) and the Allerton interests (later acquired by Morris & Co., which in turn was acquired by Armour) practically controlled the dressed beef industry. (FTC, *Report on Agricultural Income Inquiry*, p. 196.)

In 1903, the Big Five slaughtered nearly half of the number of animals slaughtered by interstate slaughterers or under Federal inspection. (*Ibid.*, p. 197.) Figures for years are as follows (*Ibid.*, Table 55, p. 198):

Year	% of all animals	% of cattle
1908	59.7	74.9
1916	70.5	82.2
1919	69.3	78.5
1924	60.6	73.2
1929	58.7	69.5
1935	66.2	67.1

Because of ownership of special facilities to distribute meats, the companies were in a position to distribute groceries with no substantial increase of overhead; by not charging the grocery division of the business its share of overhead, they were in a position to lower prices temporarily

(Caption to table reads: Big Five production of the number of animals, slaughtered by interstate slaughterers or under Federal inspection, by kinds, for specified years, 1908-1935.)

The 1917 Federal Trade Commission investigation reported that at that time there was only one independent packer who slaughtered as much as 1 per cent of interstate total of cattle, only 9 independents that slaughtered as much as 1 per cent of the interstate total of hogs. (FTC, *Report on Meat Packing Industry*, Part I, p. 33.)

In 1933 Swift, Armour, and Cudahy sold 55.5 per cent of the fresh beef production, 83.8 per cent of fresh veal production; Swift, Armour, and Wilson sold about 25 per cent of the fresh pork production, a little over a third of the cured and processed pork production, and about 23 per cent of the lard. (FTC, *Report on Agricultural Income Inquiry*, p. 194.) Of the total value of all meats, the combined sales of Swift and Armour were: 1921-55 per cent; 1929-55 per cent; 1935-61 per cent. (TNEC Monograph No. 35, p. 19.)

2. Control in related lines

A. By-products

(1) Hides. In 1916-1917, the Big Five handled more than three fourths of of the hides and skins produced by interstate slaughterers, and, directly or through subsidiaries, tanned a large part of the leather processed in the United States. (FTC, *Report on Meat Packing Industry*, Part I, p. 37.) In 1933, Swift, Armour and Cudahy sold 56.3 per cent of the hide production of the United States. (FTC, *Report on Agricultural Income Inquiry*, p. 194.) In 1935 Armour through a controlled subsidiary ranked first among the leather producers of the United States, while Swift's company was among the half dozen leaders. (*Ibid.*, p. 220.)

(2) Fertilizers. In 1917 the big packers handled more than $\frac{2}{3}$ of the animal fertilizers and 19 per cent of mixed fertilizers. (FTC, *Report on Meat Packing Industry*, Part I, p. 37.)

(3) Rendering business. The big packers had a virtual monopoly in 1917. (*Ibid.*, Part III, p. 61.)

B. Substitutes for meat products

The big packers have extended their control into many fields, handling important substitutes for their products. A few figures for illustration:

(1) Butter. In 1916 Swift was the greatest butter distributor in the United States, handling as much as the combined sales of the two largest non-packer organizations. (*Ibid.*, Part I, p. 35.) In 1934, the two largest producers of butter were Swift and Armour. (FTC, *Report on Agricultural Income Inquiry*, p. 247.)

(2) Cheese, poultry, eggs. In 1916, the Big Five handled at least half the interstate commerce in cheese. (*Ibid.*, p. 199.) In 1934 Armour, Swift, and Cudahy accounted for 36 per cent of the factory production of cheese. (*Ibid.*, p. 250.)

(3) In 1934 Armour was the sixth largest producer and distributor of canned milk. (*Ibid.*, p. 255.)

(4) The large packers in 1917 produced 31.8 per cent of the domestic production of cottonseed oil; Armour and Swift alone produced 60 per cent of the oleomargarine. (*Ibid.*, p. 198-199.) They had become important distributors of canned and cured fish, particularly salmon. (*Ibid.*)

3. Control in unrelated lines

In 1917 the packing companies were well established in the distribution of wholesale groceries, including particularly canned fruits and vegetables, and staples. In 1917 the Armour Grain Co. handled 23 per cent of all receipts of grain at Chicago. (FTC, *Report on Meat Packing Industry*, Part I, p. 37.) (The 1920 Packers' Consent Decree forced the packers to discontinue grocery distribution.)

to eliminate rivals less fortunately situated. Meanwhile their profits from other lines continued. They could always outlast independent competitors in any price war.²⁷

Manipulations of Laws Concerning Standards

Laws may be misused in many ways to exclude new entrants to an industry, or to penalize severely certain types of producers. The patent laws are sometimes used in such a way, as we have already indicated under the general heading of creating bottlenecks. Here we shall confine ourselves to a few common practices by which certain groups of producers have obtained the support of the law in excluding rivals under the guise of protecting the public against faulty products and poor work.

Unreasonable licensing requirements: an example from the plumbing industry

One of the most important areas in which laws are used as tools to support exclusive positions of private interests is the construction industry, where the restrictive practices of business are closely interwoven with those of labor. Controls over the marketing of plumbing equipment may be cited as an example:²⁸

Many building codes require, for example, that plumbing be done only by licensed plumbers, the ground for the requirement being the danger that incompetent plumbing installations will involve hazards to public health. The actual administration of the machinery for examining plumbers and granting licenses is frequently placed in the hands of local officials who are or have been licensed plumbers and who are keenly aware of the private interests of the plumbing trade. From time to time there are complaints that tricky examinations designed to limit the number of successful applicants have been used in an effort to protect established plumbers from competition. In some jurisdictions an effort is made to interpret the license requirements in such a way as to protect the master plumbers' lucrative trade in plumbing equipment from the competition of mail order houses. A Pennsylvania plumbing code, for example, forbids persons who are not registered plumbers to advertise or display plumbing equipment for sale at retail.

Unreasonable requirements concerning materials or conditions of sale

The most striking and notorious examples of unreasonable provisions concerning materials are again in the construction industry. Antiquated provisions of building codes set up before the development of new materials and processes remain on the books because of the pressures from vested interests (both labor and business) seeking to prevent the introduction of prefabricated housing and other buildings.²⁹ Mr. Edwards

²⁷ U. S. Reports, Vol. 286, *Cases Adjudged in the Supreme Court at October Term; 1931*, pp. 106-121.

²⁸ Edwards, Corwin D., "Trade Barriers Created by Business," in the *Indiana Law Journal*, December, 1940, p. 169.

²⁹ The complaints of Sears Roebuck against a provision of the building code of the city of Dayton, Ohio, is analogous to these specifications concerning the product.

provides us with an interesting illustration from quite another sphere, however:³⁰

It is alleged that the campaign of ice cream manufacturers against the use of the counter freezer has included efforts to secure the enactment of sanitary legislation for ice cream manufacture which will require the sterilization of all equipment with live steam and the use of a cement floor which slopes toward a central drain. Neither requirement is burdensome to a factory, but a soda fountain which could obtain a counter freezer only by installing a sloping cement floor, a central drain, and a steam boiler would be unlikely to make its own ice cream.

Public requirement that sellers use a privately owned grademark

In the course of its investigations, the Antitrust Division of the Department of Justice recently unearthed an interesting and very restrictive provision in the lumber industry. Members of the lumber manufacturing industry had succeeded in getting written into the standards for lumber specified by the United States Department of Commerce a provision that certain kinds of lumber be labeled with the grademark of the Trade Association. They also managed to get the government to purchase only such grademarked lumber, and in some parts of the country to require that such lumber be used in construction of houses the loans on which were guaranteed by the Federal Housing Administration. They even got such requirements written into some of the local building codes in the far West. An elaborate system was developed to limit the availability of the grademark to insiders, or to make the obtaining of the grademark very expensive for outsiders. The result was the exclusive control of sales to government, and in some areas of the entire local market, in the hands of the members of the association.³¹

Conclusion

Out of this mass of material what conclusions may we draw? It is apparent that businessmen in many lines of production and over a period of many years have attempted to gain a position of centralized control over prices and production in entire industries. They attempt to gain that control by expanding the unit of business enterprise, as in close-knit federations. They attempt to gain that control by banding together in loose-knit federations of many kinds, aiming at coordinated policies and the elimination of competitive tactics. They attempt to gain that control by focusing on those points in an industry that are most readily turned into bottlenecks for the rest of the industry, by unfair competitive tactics, by misuse of legal standards. Such policies are pursued by dominant

It is provided that a sticker must be attached to each piece of plumbing equipment; the sticker is to be obtained from the city by making weekly reports including information concerning the place of installation of each piece of equipment. A mail order house seldom has this information, and the provision therefore in effect prevents the competition of plumbing sold other than locally. (*Ibid.*, p. 173.)

³⁰ *Ibid.*, p. 172.

³¹ *Ibid.*, pp. 172-175.

businessmen representing single big business consolidations and by joint actions of those controlling leading forms in loose-knit federations alike. In some cases entrepreneurs have obtained government support in cartel-like arrangements to raise and maintain prices.

The effects of most of these practices on the functioning of the economy are dramatic and important. They block progress; those with vested interest to protect continuously erect barriers to prevent the entry of new firms and the competition of new and improved products and processes. They lead to high prices and restricted outputs; with these high prices come idle resources and waste, with lowered planes of living for the population as a whole. They deprive producers who are on the "outside" of freedom to enter the industry of their choice, and they strangle many who make the attempt. Finally, they weaken the bargaining positions of those dealing with the dominant groups, by eliminating alternatives; the tobacco farmer has no really free market in which to sell his tobacco; the buyer of aluminum ingots has no alternative source to which he can turn for his aluminum.

CHAPTER 27

Some Monopolistic Price Policies

WHEN there is only one dominant firm in an industry, it is evident that that firm will tend to use its position to set and maintain high prices of what it sells and low prices for what it buys in order to obtain profits as great as possible. When there are only a few large firms in an industry, collaboration in policies of raising or maintaining prices and restricting outputs is almost as inevitable. The reasons are readily seen; collaboration, whether tacit or formal, makes possible profits otherwise unattainable and prevents drastic price wars. The best way of gaining insight into these practices is to examine briefly a few outstanding examples. We shall therefore consider (1) some examples of extreme price policies and their effects (2) some examples of policies of price stabilization, and (3) recent price policies under governmentally sponsored cartelization.

Extreme Price Policies

Intelligent price and output policies will take into account both long- and short-run effects. It may be that for a short time greater profits would be attainable at very high prices, but in the long run this would bring new firms into the industry and profits would be less. The most extreme high price policies are therefore likely to appear: (1) when businessmen are short-sighted, (2) when they are strongly entrenched in a position to prevent the entry of new firms, or (3) when speculators are manipulating the market for their own enrichment even to the anticipated detriment of the business itself. The second of these conditions is most likely to appear in industries that are very highly centralized, this centralization both reflecting and supporting the strong walls that have been built around the exclusive positions.

Examples of extreme price policies of the big trusts of the nineteenth century

Extreme price policies have frequently been pursued by dominant concerns in periods of great economic activity, especially at the end of the nineteenth century and in the late 'twenties of the twentieth century. Despite more recent extreme price policies, however, the drama of the late nineteenth century still remains vivid. Outstanding examples at

that time were the Standard Oil Company, the American Sugar Refining Company, and the Whisky Trust. We shall look briefly into the history of the first two.

Most successful was Standard Oil, whose early attempts to exploit its dominant position and set very high prices illustrate both the strength and weaknesses of such a policy.

After an attempt at price control by voluntary cooperation among independent oil refiners in the early 'seventies broke down, Mr. Rockefeller and his associates proceeded to extend their controls through purchase and contract. By 1876 the Standard Oil Company controlled directly 90 per cent of the refining capacity of the country. Price margins between crude and refined oil were boosted from 10 cents in 1874-75 to 20 cents in 1876.¹ This high price could not be maintained. Exporters and foreign purchasers resisted the increase, new refineries were set up abroad, and consumers resorted to shale oil. The margin between crude and refined oil prices dropped to 5 cents, but Standard continued buying up refineries until in 1880 it controlled 95 per cent of the industry. It then boosted prices again. The reaction was even stronger and foreign and domestic competition was stimulated. Prices were quickly dropped again. Standard could keep prices well above costs, but there were clearly limits as to how high they could go—limits in the shape of the demand curves for refined oils and in the potential competition of both foreign and domestic producers. Standard officials in the future recognized these limitations and sought to maintain high but not such extreme prices. Miss Tarbell quotes Benjamin Brewster's evidence before a federal commission investigating the Standard.² Asked if the company could set prices at will, Mr. Brewster replied, "At the moment many things can be done, but the reaction is like a relapse of typhoid fever. The Standard Oil Company can never afford to sell goods dear. The people would go to dipping tallow candles in the old-fashioned way if we got the price too high."

The sugar trust was less successful than the Standard Oil Company in maintaining a position of leadership and high prices. There were several periods in its early history, however, when the trust was successful in setting prices high enough to obtain extremely high profits even on inflated cost figures. Jones summarizes the history of sugar prices during the years in which the American Sugar Refining Company held its dominant position in the industry:³

The effect of trusts on prices is shown in illuminating fashion by the experience of the sugar trust . . . it appears that the margin between the price of raw sugar and of refined sugar was high during the early 'eighties, and declined rapidly after 1882 and until 1887 (the year in which the sugar "trust" was

¹ Tarbell, Ida, *History of the Standard Oil Company*, Macmillan, New York, 1933, Vol. II, p. 201.

² *Ibid.*, Vol. II, p. 206.

³ Jones, Eliot, *The Trust Problem in the United States*, The Macmillan Co., New York, 1921, pp. 262-263.

formed). The decline in the margin between 1882 and 1887 reflected the keen competition that prevailed—a competition so severe that only those refiners who realized the economies of large-scale production were able to operate at a profit. Many refiners, particularly those who failed to envisage the inevitable trend toward larger production units, were indeed obliged to withdraw permanently from business. In October, 1887, the “trust agreement” became effective; and the margin rose from three-quarters of a cent per pound (in 1887) to one and one-quarter cents (in 1888), an increase of approximately 65 per cent. No doubt the margin was abnormally low prior to the formation of the “trust”; and therefore it is difficult to say how much of the increase is fairly attributable to it. The high margin of 1888, however, speedily attracted new competition; and as a result the margin fell in 1890 to an even lower figure than during the ‘eighties. In 1892 the trust, through the acquisition of a number of competitors, secured nearly a complete monopoly of the sugar refining industry; and the margin was considerably advanced once more. As before, this induced new competition, as the result of which the margin fell below the cost of refining. Upon the acquisition of several competitors in 1900, prices and margin again went up; but this led to the construction of competing refineries, and in 1904 the margin again declined. Taking, therefore, the first eighteen years of the life of the trust—the margin after 1905 indicates the existence of competitive conditions—it appears that sugar prices were low when competition was present, and were advanced when competition was absent or brought under control.

Other examples of high-price policy by a dominant leader might be cited. The control may be more or less complete; but in most cases extremely high prices are broken by the resistance of buyers, and usually also by the appearance of competing enterprises and competing goods. In some cases it has proved possible to maintain moderately high prices (above costs) for long periods, but to do this effectively the dominant firm must have enough props to ensure continuance of its position in the market.

An example of recent high-price policy

An outstanding example of high-price policies in the active days of the ‘twenties is afforded by the anthracite coal industry. All of the anthracite mines are located in a limited area in Northeastern Pennsylvania, and control over this property was early established by the great eastern railroads. The Hepburn Act, passed in 1906, contained a “commodities clause” that prohibited the railroads from transporting in interstate commerce goods that they produced and owned. By this provision it was hoped that the monopolistic control of anthracite production would be broken. There has, however, been continuous evidence that competition has not been operating freely in this market. Although the railroads no longer own the mines outright, they have maintained effective control through linkages with financial interests and through interlocking directorates. During the ‘twenties they used their power to raise prices on anthracite coal to higher and higher levels; and prices were not permitted to drop during depression periods. In the years 1917 to 1929 production dropped off a fourth and price was nevertheless raised by about the same proportion.⁴ The mine operators continued with this high-price policy despite the increasing competition of other fuels: oil,

⁴ TNEC Monograph No. 21, p. 181.

gas, and bituminous coal. Consumers shifted in increasing numbers to the use of these other fuels. There were many evasions of the price quotations, and an increasing shift of trade from the "line companies" (those under the influence of the railroads, and previously belonging to the railroads) to the "independents." A bootleg trade developed, carried on in part by unemployed miners who siphoned the coal off of company property and shipped it to the cities by truck. The operators finally managed to get the support of the governor of Pennsylvania in the development of a program to share the market by planned curtailing of production in which all producers ("line" or otherwise) were expected to participate. This policy, developed formally in 1939 and 1940, served to support high prices despite the weak position in which the anthracite operators found themselves.⁵ It was possible only because the limited area from which anthracite coal could be mined automatically limited somewhat the possibilities for entry of new enterprises in response to high prices.

Speculative manipulation of prices by outside financiers

When there are only a few dominant firms in an industry, it may be possible for those who have obtained financial control to manipulate prices in order to speculate in the stocks of the controlled enterprises, or to profit on investment in these stocks over a fairly extended time period. A striking example is the sharp advance of copper prices and their pegging in 1929. "During no year from 1921 to 1927 inclusive had the price of copper averaged higher than 14.5 cents. For the year 1927 it was 13 cents a pound. . . . In 1926 and 1927 . . . [the] rapidly growing concentration of ownership of [copper-producing properties was] supplemented by the formation of two industry-wide organizations, Copper Exporters, Inc., and the Copper Institute. Building on this foundation it was possible in 1927 to weld the entire copper industry of the United States and the world into one entity so far as price-making was concerned. No sooner had this unity been achieved than prices began to move up. As against a 13-cent average for 1927, the figure for 1928 rose from 13.9 cents in January to 15.8 cents in November and December. By March 1929, it

⁵ The production control scheme introduced by the leading producers in 1939 and the stronger program undertaken in 1940 are both very simple. The 1939 plan involved an agreement to limit the number of days per week that the collieries should operate. This provided an automatic sharing of the market according to the capacity of the mines. The program was enforced largely through moral suasion with the support of the governor of Pennsylvania. The 1940 plan is administered by a committee selected by the governor and composed of nine members—three representatives of the governor, three representatives of the operators, and three representatives of the miners. A board made up of representatives of the operators determines what should be the total output of anthracite, and then this amount is allocated to the different companies according to their production in the two or three years prior to 1940. There is no price-fixing agreement, but there is moral suasion to comply with the output quotas and to refrain from price competition. The really strong arm of the agreement comes from the participation of the United Mine Workers in it; this labor organization can bring any offending operator in line very easily by making things difficult for him in the labor market.

reached 24 cents. Sales of copper abroad had, however, begun to fall off as the price passed 14 cents; and the copper men themselves recognized that a price of 24 cents was too high. It was, they claimed, forced on them by a market that anticipated a shortage. However this may be, from April 1929 through March 1930, not budging one iota in the face of general business collapse, they pegged the price of copper at 17.8 cents a pound.

"In April 1930 the price broke and by October it was under 10 cents. . . . In December 1932 it reached an all-time low of 4.8 cents, from which it climbed slowly back to 9.5 cents in 1936."⁶ The raising of the price of copper and the pegging of that price at such high levels at this particular time represented "not only a disruptive but an almost suicidal policy on the part of those who developed and exercised the power to determine copper prices." The price boost came just as new mines were developing in the other countries; it encouraged the opening up of high-cost mines in the United States, and the price rise was in the face of drastic drops in the prices of other goods.

The Brookings Institution study which we have quoted points to the strong evidence that the maintenance of such extreme prices was due to the control of policy by bankers manipulating copper stocks, both to expand the Anaconda Copper Company and to gain speculative profits. The price rise was far greater than that which, apart from financing operations, would have been charged by a simple monopoly in complete control of the market. The relapse afterward was similarly extreme.

Price Stabilization as a Business Policy

Relatively stable and moderate price policies are likely to be found: (1) Where there has been an experience of violent price cutting and new competition that firms now wish to avoid, (2) when the practical difficulties of getting effective collaboration among firms are too great if price changes are frequent (as is to be expected when policies are extreme), and (3) when there is considerable likelihood that extremely high-price policies would lead to serious evasions and the breakdown of collaborative pricing by firms dissatisfied with the curtailment of sales involved, and (4) when the dominant firms lack strong tools to prevent the entry of new firms. The steel industry provides the outstanding example of a long period of collaborative pricing, initiated by a dominant price leader (United States Steel) in which policy has definitely been directed to the stabilization of prices through good times and bad.

Incentives to collaborative price stabilization in high-fixed-cost industries

Before turning our attention to the history of price policies in the steel industry, we should consider briefly what would happen (and has

⁶ Nourse, Edwin Griswold, and Drury, Horace B., *Industrial Price Policies and Economic Progress*, The Brookings Institution, 1938, pp. 150-151.

happened) in high-fixed-cost industries when firms have engaged in very aggressive price wars during depressed periods.

An industry in which relatively fixed costs are high and fixed agents very durable is peculiarly susceptible to the vicissitudes of "outrageous fortune." Since the fixed agents wear out slowly, a firm once in the industry will hang on through years of poor business, operating at a loss. This problem is exaggerated when equipment is highly specialized so that it is difficult or impossible to shift it to other uses. When "bad times" come to an industry of this kind, each enterprise finds its sales curtailed. With the smaller sales, fixed costs per unit are higher and there is a strong incentive to cut prices in order to expand output to a more profitable level. A concern seeking to expand its output then cuts prices below those charged by its rivals. As it draws customers away from rivals they respond by cutting their prices in turn and so the process goes on. If demand for the product is very elastic, all may have considerably larger sales at the new lower price, but if demand is quite inelastic the increase in sales of all firms will be very small. Each firm acting alone finds the demand for its product almost (though not quite) perfectly elastic and the marginal revenue immediately resulting from an increase in output at the slight price reduction needed is therefore almost (though not quite) as great as the new selling price. Each firm acts almost as under pure competition. As a result, these "cut-throat" tactics will bring prices down approximately to the level they would reach in the short run under pure competition. With high fixed costs this short run may last a very long time.

During the bad times equipment is allowed to fall into disrepair, some is scrapped, and eventually some firms drop out of the industry. As this dropping out takes place, prices rise again. The price rise is sometimes exaggerated by changed conditions in demand (as was the price fall). For a while profits in the industry may be very high. It takes some time to get the wheels of the industry running smoothly. New firms appear, but it takes a year or more before their full production capacity becomes evident in the flow of goods onto the market. During this interval the industry is quite likely to be overexpanded. As the new firms are completed the market is flooded and price cutting sets in again. Thus both prices and output are very unstable. Short-run prices essentially competitive fluctuate violently to each side of the long-run, purely competitive level.

Manufacturing industries with high fixed costs rarely have enough sellers to be purely competitive. The violent ups and downs are experienced for a while, but they are an incentive to both formal and tacit agreements to keep prices up. The fewer the sellers, the easier it will usually be to accomplish some type of "cooperative" price maintenance. This may be imposed directly or indirectly by a leader firm. Trade associations may be used directly or indirectly for this purpose, appeals may be made to government as was the case during the National Industrial Recovery Administration program. Once accomplished, this "coopera-

tion" will often be carried beyond the maintaining of cost prices over a period of time to the maintaining of prices at a level well above total unit costs, at the operating level of most or all firms. In an industry with a history of violent price fluctuations and cut-throat competition, a tradition of restraint may develop even without direct "cooperative" action. One among a few sellers of a good cannot reasonably ignore the effects of his price policy on the policies of his rival. In any concrete situation he will consider what such repercussions may be. Once a price-cutting war is under way, he will have to retaliate in self-defense, but each producer will hesitate to initiate a price reduction in a relatively stable situation unless the pressure to do so is very great. Prices as high as those that would prevail if the entire industry were controlled by one firm would rarely if ever be long maintained when such maintenance rested solely on individual oligopolists refraining from price cutting. Even if existing firms were to maintain prices at such a level, if it were far above costs and if profits were high new firms would inevitably appear and bid for sales at lower prices. As prices near cost levels, however, the members of the industry become more and more reluctant to incur the hostility and retaliatory price cutting of rivals. A price above costs may be maintained for long periods when members of an industry see it to be to their ultimate advantage to refrain from price cutting, and yet those contemplating entrance into the industry face the potential of cut-throat competition to keep them out should they make the attempt.

Price leadership and stabilization of steel prices

Price maintenance usually means in practice the stabilizing of price over more or less extended periods. The industry most frequently invoked to illustrate the effects of a policy of price stabilization by a price leader is steel. This is an industry which has at times suffered the sharp fluctuations in prices associated with high fixed costs and cut-throat competition. The fact that the demand for many steel products is very inelastic and very fluctuating over time increases the instability inherent in uncontrolled rivalry between high-fixed-cost enterprises. The optimum scale of enterprise in most sectors of the industry is large and firms are few in number. It is, therefore, relatively easy to arrive at some sort of "cooperative" agreement, and the incentive to do so is unusually strong.

United States Steel acted as leader in setting prices of steel products over many years. This was accomplished at one time through formal pooling agreements, later through the famous "Gary dinners" and the "gentlemen's agreement," unified policies being agreed upon without any formal contract. The Gary dinners of 1907-1911 are generally considered to have been very important in "inculcating into members of the industry a realization of the benefits of leadership."⁷ The price leadership of the United States Steel in 1911 was evidenced by the almost unani-

⁷ Burns, Arthur Robert, *The Decline of Competition*, McGraw-Hill, New York & London, 1936, p. 79.

mous agreement of other manufacturers at a meeting at which Judge Gary, President of United States Steel, stated that he believed it would be unwise to lower prices at that time. "Unreasonable" rivals were brought into line by threats of retaliatory price cutting if they offered goods at less than the agreed price. Rival firms have grown in relative importance, especially since 1920. "In 1928 it was the president of the Corporation's [United States Steel's] largest rival who was telling members of the American Iron and Steel Institute that 'the avoidance of uneconomic price cutting' was necessary to the stabilization of the industry"; and in 1932 producers appeared to be inclined "to turn more to the American Iron and Steel Institute as a means to the control of competitive conditions."⁸

To keep prices stable was the policy emphasized by Judge Gary and his successors, and is still the chief policy of the American Iron and Steel Institute. Controls by the United States Steel corporation were never perfect, but for many years this policy has nevertheless prevented violent fluctuations in the prices of products of a high-fixed-cost industry. The reasons for pursuing a policy of stabilization rather than making frequent readjustment of prices to changing conditions of demand are to be found in the long-range views of outstanding steel men who first established the tradition, in the conditions of the industry that made it peculiarly susceptible to the devastations of cut-throat competition in bad times, and in the practical fact that stable prices are much more easily enforced by a leader than are frequently shifting prices.

Even a policy of price stabilization has become increasingly difficult to enforce. For example, economic pressure for quantity discounts to big buyers is particularly strong in an industry in which large orders cost much less per ton than do small ones. Some of the anarchy which lies in back of the apparent uniformity of steel pricing is described lucidly in an article in *Fortune Magazine*.⁹ "There are, in effect, two price systems in steel. One is open, public, unyielding. This is the bleak prospect that the average consumer faces. But the big consumers buy their steel under another system, an unofficial and undercover business of concessions from the published . . . prices. Under this system the base price is merely a mark to be shot at, and the only question is how much it will be 'shaded.' These more fortunate buyers are those who possess some form of bargaining power, reciprocal buying, strong financial connections, personal influence, and, most important of all, the ability to place big orders. How much steel is sold at shaded prices is a question—since secrecy is the essence of the whole business. Undoubtedly the practice is widespread."

The curtailing of demand in the recent depression added pressures on steel producers to cut prices below the announced rates. The United States Steel talked loud and hard for "cooperation" in maintaining an-

⁸ *Ibid.*, p. 81.

⁹ "U. S. Steel II: Prices," April 1936, pp. 132 and 134. Copyright TIME, Inc., 1936.

nounced rates, and even sought to meet the situation by increasing rates, but many "independent" producers proceeded without talk quietly to cut prices. For the price-maintenance group, the New Deal provided a rescue in the National Recovery Administration. Commenting on this, the editors of *Fortune* state: ¹⁰ "Under the Steel Code for the first time the industry was able to enforce legally—or at least quasi-legally—a far-reaching ban on price cutting. Under the code, any producer who wanted to sell his steel below the published base price had to report his new price at once to the code authorities, who immediately published it. Violations carried a penalty of \$10 a ton. To rob price cutting of its remaining charms, the code furthermore provided that a cut could not be put into effect until ten days after it was publicly announced and that any producer who cut his price in consideration of a customer's promise to give him a specific order after the ten-day period elapsed was also guilty of a code violation, at \$10 a ton." The unity established under the NRA survived the demise of that experiment, but pressure from big buyers, especially from Henry Ford, has increased, and has apparently brought some success in pushing down the announced prices of sheet and strip steel. Mr. Ford has been less successful in exerting a downward pressure on the prices of iron ore.

That the tactics of the United States Steel and the American Iron and Steel Institute have generally pegged steel prices at levels on the average exploitative over extended time periods is evidenced by the small percentages of capacity generally in use, and the excess capacity built. Even high profits and restricted output do not prove that the public is worse off than it would have been with uncontrolled price cutting in bad times and high prices in good times; but high average profits do mean that stabilization has been at a price level higher and an output smaller than that which would be consistent with the welfare of the general public. During recent years there seems to be little doubt that the United States Steel and its followers in the Institute have assumed the demand for steel products to be very inelastic, that is the effect of price on amount taken has been assumed to be negligible; they have therefore sought to establish high prices even in the face of small demand and much idle capacity.

Basing-point pricing ¹¹

An important tool of price maintenance policy in some industries in which transportation costs are high is the basing-point system of pricing. It has been highly developed in steel and some other metal products, in lumber, sugar, and especially cement. Under this scheme prices are uniformly quoted at any given place as the established basing-point price

¹⁰ *Ibid.*, p. 134. Copyright TIME, Inc., 1936.

¹¹ The basing-point system is only one of a whole set of pricing schemes designed to support price maintenance agreements (tacit or formal). Others are the zone-price system, and the freight-equalization plan.

plus transportation charges from the basing point, whether or not the product actually comes from that point. Copper, for example, is sold from smelters in Arizona, Montana, Michigan, New Mexico, California, and Tennessee on a New York basis.

The most famous example of basing-point pricing has been in the steel industry. By the end of the nineteenth century, prices on most steel products were based on Pittsburgh prices plus transportation costs from Pittsburgh to the point of delivery. United States Steel early established the "Pittsburgh plus" basing-point system, under which all steel producers quoted as uniform prices the established market price at Pittsburgh plus transportation charges from that point to the point of delivery. Thus the auto-part manufacturer in Chicago, even though his steel came from a near-by steel mill only a few blocks away, was quoted the Pittsburgh price plus transportation charges to Chicago. The effect on buyers was discrimination against those furthest from the Pittsburgh area.

Since its establishment the Pittsburgh plus system has been greatly modified, partly as a result of the pressure of growing competition within the industry but probably more because of legal pressures. In 1907 Birmingham was made a separate basing point for certain products. Later Chicago was added, and in the last two decades the list has been expanded to include a considerable number of the more important steel producing centers. Of course, as quoted prices have been shaded by one means or another, the apparent fixity of the basing-point system has been undermined. None the less, the geographical pattern of discriminatory prices on many steel products rests today on a modified basing-point system.

The advantages of a basing-point system from the point of view of the dominant steel companies is evident, and the effects of the basing-point system as a means of maintaining stable prices have been emphasized by steel producers themselves. By the use of this system it was possible to get uniform published prices (calculated on the price at the nearest basing point plus freight rates) at each point of delivery of steel. Both intentional and unintentional price cutting could in this way be more easily avoided. Basing-point pricing was also a tool that could be manipulated, at least temporarily, to the advantage of a dominant group seeking to maintain its position in the industry. Thus in the early days of the Pittsburgh plus plan, Pittsburgh producers could compete with producers anywhere at equal prices so long as these other producers followed the leadership. They followed the leaders when prices set were satisfactory to them or under the threat of price wars. Finally, the effects of the basing-point system depend on the extent to which announced prices are the actual prices. When special buyers get quantity discounts these discounts are not likely to be systematically related to costs. When base prices are in fact maximum rates charged only to weak buyers or to the United States Government (as was indicated by Mr. Grace's evidence before the Temporary National Economic Com-

mittee),¹² they exaggerate price discrimination under the illusory veil of price uniformity.

The extent of price stabilization

The steel industry has set a pattern of price leadership and price stabilization that has been followed in many other instances, whether formally or informally. Outstanding examples are agricultural implements, petroleum, anthracite coal, cement, tin cans, tobacco. Efforts to maintain prices have been widespread.

The history of commodity price movements shows both the strength and weakness of such arrangements. In a large number of important cases quoted prices show an amazing degree of stability over long periods. Taking Bureau of Labor Statistics wholesale price figures from 1926 through 1933, a period when price changes would be expected to be most rapid, 73 out of the total of 784 prices changed four times or less over the seven years. Of the 73 about half changed two times or less. To take an extreme case, the list price of oil drilling bits of a standard variety has remained unchanged from 1894 to the present. Allowing for some bad selection and reporting of prices, these figures are still an impressive testament to the prevalence of price maintenance. In analyzing the individual prices in the group, a large number are found to be outstanding cases of known or strongly suspected price-maintenance agreements. These and other similar data indicate the force of price-control arrangements.

However, it would be a serious error to suppose that the remarkable stability of list prices proves the success of the industries in preventing price competition over the period. In such cases there are a variety of means of competing for business without changing the list price. In many industries substantial discounts from the list are standard practice, especially for large buyers, and these discounts can be varied at will without affecting the list price. Concessions may be made in credit terms. Delivery costs may be absorbed. Trade-in allowances may be increased. Secret rebates or other concessions may be given. Aside from all these possibilities is the entire realm of variations in the quality of the product. In many cases price competition has been confined within these limits; while in other cases the evidence indicates that price shading by various means almost always precedes cutting of list prices.

Recent Price Policy Under Governmentally Sponsored Cartelization

The participation of government in price making has developed on two very different fronts. The first is a kind of price regulation which is of long standing; it is the setting of maximum prices on the goods and services produced by monopolies such as the local electric light and power

¹² Hearings before the Temporary National Economic Committee, Part 19, 1939, pp. 10594-10600.

companies or the railroads. Regulation of these "public utilities" will be discussed in Chapter 29. The second and relatively new front is the supporting or fixing of minimum prices where for some reason these prices were considered to be "too low" in the absence of government intervention. Barring government collaboration in the NRA, this support of minimum prices has been primarily in agricultural industries and in the bituminous coal industry. What has been the performance of these industries and of government in recent pricing practices?

Pricing under the Agricultural Adjustment Administration

A full evaluation of the government's program in agriculture cannot be undertaken until our examination of the total economy is much more complete, and an over-all picture of problems and policies in agriculture has therefore been postponed to a special chapter at the end of this book. A few brief comments, however, are essential as part of a study of collaborative pricing in the American economy.

There have been two major stages in the recent history of price policy in agriculture in the United States. The first of these was the attempt to maintain prices that would bring adequate incomes to farmers in a depressed agricultural situation. This program was justified primarily on the grounds of claims that farmers were suffering unduly in the existing economic situation, that their incomes were low and that in the face of restrictive practices elsewhere in the economy they needed public assistance for the raising of agricultural products. The chief criticisms of this government sponsored program were (1) that it tended to help the high rather than the low income farmers, sometimes actually penalizing those at the lowest income levels, (2) that it failed to accomplish the long-run goal of moving excess resources out of agriculture and into other productive employments, and (3) that the prices set were too high (they were based on a "parity" concept which implied making them as high relative to other prices as they had been in the period of greatest agricultural prosperity of this century). The second phase of governmental protection of minimum prices in agriculture has developed in the war situation; guarantees of minimum prices to farmers have been undertaken in order to provide adequate incentives for the expansion of agricultural outputs to meet war needs. Experience here is still inadequate to justify clear-cut conclusions; but it is evident that the present need for expansion introduces a major problem concerning future agricultural policy. Will a postwar readjustment in agriculture emphasize consumer needs and a broad nutritional program, with subsidies to agriculture? Will it return to a restrictive policy designed to curtail production and thus maintain prices? And if restriction is the policy decided upon, will the program be developed in such a way as to facilitate the shifting of resources out of agriculture? These are questions which only the future can answer.¹²

¹² Dr. Margaret Reid comments on the implications of a program guaranteeing agricultural prices: (*Consumers and the Market*, 3rd ed., Crofts, 1942, pp. 541-542) "The desirability of using guaranteed minimum prices depends also on their use when

The pricing of fluid milk

Government participation in the control of milk prices has been both federal and state. Federal control has been undertaken through marketing agreements in some areas; these agreements may involve any or all groups in an area—processors, distributors, cooperatives. Local controls have been through state milk control boards. In some cases only prices paid farmers have been controlled; in others, both farmers' prices and retail prices have been fixed.

The goal set in federal control of milk prices has been essentially that sought under the Agricultural Adjustment Administration in general; the attempt was to get for fluid milk "parity" prices that would be in the same ratio to other prices as the ratio prevailing in a "base period" that was highly profitable for the dairy industry. In the Agricultural Marketing Agreement Act of 1937, the Secretary of Agriculture was empowered to plan for rates even higher than this if it seemed "reasonable" to do so, taking into account costs and other relevant factors. It has not proved possible to set prices as high as the attainment of these goals would require. A group of economists studying the industry¹⁴ have observed that:

Competition of fluid milk from other markets, unregulated by Federal authorities because interstate commerce was not involved, and competition from producers whose milk was formerly sold largely for manufacture of butter, cream, and other manufactured products have imposed practical limits on the level of fluid milk prices that kept them below parity. *Consequently the objective seems to have been to fix the highest milk and cream prices in each market that could be maintained successfully.*

The same attempt to fix prices at the most profitable level possible has also characterized most of the price fixing under state controls. Thus the same authors write:¹⁵

In Wisconsin "the authorities have apparently set prices as high in relation to this goal (cost of production)¹⁶ as they thought could practicably be maintained

it seems desirable to contract production. It might be necessary to lower the minimum price in order to discourage further accumulation of reserves. Resistance to this is likely to occur, and an effort may be made to maintain high guaranteed prices and to expand surplus disposal programs. Subsidies to prevent a fall in price have some merit in an emergency while resources are being shifted to some other use. They should however not be used to prevent the shift, unless in the carefully considered judgment of experts consumer interests, apart from demand being expressed in the market, require that production should not be contracted. Subsidy as a long-run part of directing production should be appraised in terms of consumption needs.

"In this as in other cases of State-administered prices, political pressures from sellers to be benefited, may become a major obstacle to sound social policy, even to the extent of suppressing the findings of Federal research bureaus."

¹⁴ TNEC Monograph No. 32, p. XXIII. (Italics supplied.)

¹⁵ *Ibid.*

¹⁶ The estimates of costs are generally too high, as is evidenced by the active competition of new producers at prices even below the estimated "cost" price.

with whatever control of sales volume the local cooperatives were able to achieve." In Oregon "prices of fluid milk are apparently set at the most profitable level, given rigid control of entry to the market and a scheme of payments to producers of fluid milk according to sales quotas that discourages expansion of production for the fluid milk market." In Indiana "apparently the administrative agency follows in large measure a policy of enforcing the prices determined by bargaining between cooperatives and distributors, limiting entry to fluid milk production whenever there is a danger that new entrants would render these prices difficult to maintain."

Although the policies pursued in the other two states studied, New York and California, seem to have been more moderate and to have included greater attention to efficiency in distribution, in all five states the results have been a boosting of milk prices over those which would otherwise have prevailed. In some cases differentials of price between home and store delivery have been abolished in order to protect the existing system of distribution,¹⁷ thus causing waste and inefficiency. The result has been both discouragement of the consumption of milk, and a high cost of milk to consumers. Moreover, uniform retail pricing of milk has worked against the small producers who could not afford to attract customers by heavy advertising expenditures and who were deprived of the alternative of charging a slightly lower price. At the same time it has encouraged the development of excess and under-used capacity in the industry.¹⁸

The pricing of bituminous coal

Ever since the end of the First World War the bituminous coal industry has been sick. It is an industry of many firms, large excess capacity, and a large quantity of excess immobile and highly specialized labor. The Bituminous Coal Act of 1937 provides for minimum price fixing intended to protect the incomes of operators and laborers, and maximum price fixing if this should prove necessary to protect consumers (primarily in time of war). Standards for the setting of price are vague, though they are supposed to be based on "costs." Estimates have been based on past cost and consumption figures so that no account is taken of demand elasticity and the expanded consumption and lower unit costs that would follow from lower prices.

Before the war price fixing involved the setting of minimum rates only; no maximum levels were established. The pricing problem is more complex than in the case of milk, because many different kinds and grades of coal are involved, because of the importance of transportation costs, because coal is sold in a national instead of in local markets, and because

¹⁷ *Ibid.*, p. XXIV.

¹⁸ It has been argued that governmental support of prices to farmers is necessary to protect farmers in selling to oligopsony groups among milk distributors. Commenting upon this Dr. Reid suggests that "the answer may be fixing of prices to farmers with prohibition on restriction of supply, plus active enforcement of anti-trust legislation. If prices are fixed unnecessarily high, expansion in supply will serve to bring about a readjustment in price." (*Consumers and the Market*, 3rd ed., Crofts, 1942, p. 546.)

of the close competition with other fuels. Problems arising in connection with price regulation of coal have already led to suggestions that other related prices should be controlled also, in particular that prices of gas and oil should be brought under control, and that the coal situation should be handled jointly with transportation policy. Each new step leads to new ramifications and to suggestions for yet further extension of government regulation. The principal criticisms of the Coal Act may be summarized briefly as follows: ¹⁹

1. The rules permit hidden profits to be included in the costs on which the minimum prices are based. . . .
2. Certain regulations of the Commission interfere with low-cost transportation. For example, they require that higher prices be charged for the same coal if it moves by truck than if it moves by railway. Furthermore basing-point pricing is allowed. It has already been pointed out that this causes cross-hauling, and is one of the practices the F.T.C. and the T.N.E.C. have recommended prohibited.
3. Nothing has been done to check a rise in selling costs that are included in the costs for determining minimum prices. With price competition checked non-price competition is likely to increase. (It is easy to write up costs in payments to sales agents, and so on.)
4. Regulation involves lags in adjustment and this tends to result in too high prices in periods of expanding demand because there is a failure to adjust immediately for the lower unit costs associated with increased output.²⁰
5. In periods of diminishing demand reduced output raises unit costs and therefore under the cost basis of price regulation tends to lead to a raising of price. This exaggerates depressive tendencies and is especially serious for the total of the economy since the coal industry is basic over a wide range of industry.²⁰
6. High prices retard the withdrawal of excess capacity from the industry and may even lead mines which had been closed down to re-open though mines already in operation are functioning at only a small fraction of capacity.²¹

All of these criticisms seem to add up to a serious indictment of the Coal Act. It has all the unfortunate features of the collusion found in private oligopoly. Yet it does not follow that there is not a case for government intervention in the coal industry in order to protect those workers and operators who get caught in the industry in a period when resources should be shifted out and into other areas of production. The criticisms are largely focused on the lack of a long-run view in making adjustments, and the tendency to follow a pattern of short-sighted oligopoly pricing in the formulation of policy and the administration of the act.

A Historical Perspective on Price Policies

Privately controlled pricing programs

Throughout our history there have been restraints of trade and unfair competition in many spheres of business activity. Formal and informal

¹⁹ The first three points are quoted verbatim from Reid, *op. cit.*, p. 549.

²⁰ See TNEC Monograph No. 32, pp. 316-317.

²¹ Reid, *op. cit.*

agreements to maintain prices and restrict outputs have come and gone with the rising and falling tides of economic prosperity and with the impact of wars on economic organization.

The extreme early price policies of some big business stand out dramatically. Prices were cut drastically and far below competitive levels in order to eliminate rivals. When the field was clear, prices were raised to very high levels. The high price was sometimes that at which profits would be maximized in the short run, but high prices often brought new firms on the scene, and another price war then ensued. In some cases prices were raised to absurd heights on the basis of faulty estimates of the elasticity of demand and a short-sighted failure to recognize potential competition. At the end of the last century extreme policies could sometimes be successful because of the rapidly expanding markets and the unfair practices used to bolster a favored position. Under these circumstances price fluctuations were truly amazing. Examples that we cited were dominant firms or trusts such as the Standard Oil, the American Sugar Refining Company and the Whisky Trust. Extreme policies were pursued not only by single dominant firms but also by pooling agreements in which several major companies were involved. Public opinion grew hotter and hotter against the new giants and their tactics. The big "trust-busting" era was on.

Experience and maturity, and a longer range vision that considered both economic and legal reactions, led to the policy of stabilizing prices—a policy pursued, for example, by United States Steel and price leaders in other industries. Experience of cut-throat competition acted as an incentive to "cooperative" action in maintaining and stabilizing price, whether set by a leader or maintained by convention in a more evenly balanced group of enterprises. Prices of anthracite coal, cotton thread, newsprint, soda crackers, and a number of other commodities have shown considerable stability for extended periods before and since the First World War (1914-18). But stable price policies may still be high-price policies. Failure to recognize demand elasticity is still a common error. Violent price wars have not been eliminated. And the speculations of promoters, as in the copper industry, still add to the turmoil.

Recent government support of collaborative pricing

Maladjustments in our economy in the 'twenties were followed by the deep depression of the early 'thirties, which finally knocked the bottom out of the markets for many goods. Some industries with high fixed costs had become chronically ill. Under these circumstances agreements weakened, but government became more sympathetic toward "cooperative" action of the members of an industry in their attempts to control prices and production. The abortive NRA strengthened these attempts and left the laws against combinations under a haze that required new interpretations. Trade associations were expanded to include larger proportions of the firms in an industry, and their activities were broadened and supported by government sanction. The anthracite coal agreement

(essentially a pool to maintain price and restrict output) was upheld by the Supreme Court. The AAA was created; the Agricultural Marketing Act was passed; the Bituminous Coal Commission came into being. Increased state intervention in pricing accompanied in some degree the expansion of federal action. Policies implementing the traditional efforts to preserve competition and keep government out of business gave way to government support, in selected industries, of price maintenance policies otherwise illegal under the antitrust laws. More government in business meant also more business influence in government action designed to control business practices.

CHAPTER 28

Public Policy Attacking Restraints of Trade in Business

PUBLIC policy directed against "practices in restraint of trade" has a long history that goes back at least as far as the thirteenth century in England. In this country since early days the courts have under certain circumstances opposed "practices in restraint of trade" as being contrary to the public interest. During the nineteenth century "restraint of trade" became closely identified in America with "unfair competition" and with "monopolistic intent." The movement against such practices, primarily expressed through the "antitrust laws," was a movement to preserve free competitive action in the economy. It took on major importance in the 'nineties when large corporations and combinations were coming into prominence; it has proceeded in irregular spurts since then and has received especial attention in the past few years.

Mr. Edwards of the Antitrust Division of the Department of Justice states in simple and telling language the consumer's case for the preservation of competition between rival enterprises:

An economy, like a political state, can avoid abuse only through a system of checks and balances. The profit motive is like the steam in an engine. Well directed, it drives the machine, but badly directed it scalds or blows up the machinists. There is everything to gain by encouraging businessmen to make more goods, to make better goods, and to find new processes. But we must discourage the type of economic activity which creates obstacles to trade in order to collect toll for letting people go through. The easiest way to discourage such anti-social activities is to expose the business which engages in them to the competition of other and more useful enterprises. In the early days of the automobile a few farmers maintained well-watered mud holes in order to get an income by towing motorists out, and the motorist's quickest protection was to find an alternate route. From the consumer's point of view, competition is primarily a device for giving the buyer of goods who encounters high prices or bad goods an alternate route to his purchases.¹

By keeping open the channels of trade, by cracking down on restraints and exclusions, competitive forces are brought into play that encourage economic progress and abundant production with full use of resources.

¹ Edwards, Corwin D., *Consumers and Monopoly Problems*, Address before the National Conference on Consumer Education, Stephens College, April 3, 1940. Mimeograph, p. 2.

But public action against restraints of trade is much more than a policy to protect consumers and increase production. It is a particular kind of policy rooted in a free enterprise system. Where Germany has encouraged cartels and then regulated them, the American approach has been to prevent the building of monopolies and to declare illegal cartel agreements that restrain trade, attempting to prevent their occurrence and to break them up when they appear. The attack on monopoly and on restraints of trade in America is an expression of the whole social political economic atmosphere and the ideology of the culture. Individual freedom and equality of opportunity are vital parts of that ideology. Industry should, from this point of view, be controlled by a minimum of interference with individual action to the end of maximizing general welfare. This broad goal requires that some individuals should not be permitted to engage in practices that seriously hamper the freedom and opportunities of other individuals, whether as consumers or producers. Actions that restrain trade interfere with both the full expression of consumer preferences in the allocation of resources and the freedom of individuals to engage in those productive activities that they may prefer. Moreover, concentration of power in industry exaggerates the inequalities of income distribution and increases the impact of big business pressures in politics. The concentration of wealth in a few hands and the concentration of power in those same few hands have evoked excited criticism periodically in our history, and most dramatically prior to the first federal antitrust legislation in 1890.

The purpose of this chapter is to trace the history and to evaluate the possibilities of public action (in the United States) to prevent restraints of trade and to tear them down where they have appeared.

Common Law and Restraints of Trade

Common law is law that is not written into statutes but is handed down through judicial precedent in court decisions. The law as enunciated in such court decisions has been termed "common" law because it is based on common customs and practices as contrasted to statute law, which has been written into law by the acts of a legislative body. American common law has developed differently in each state, but the roots of modern common law here were in the common law of England. In both England and America the courts rely on such common law (primarily as expressed in earlier court decisions) for their decisions in cases where there is no statute covering the point at issue, and also in the interpretation of the meaning of statute law. Common law was the earliest basis of legal control over practices "in restraint of trade." The cases in which such practices were questioned came up in state courts, since control over these cases had not in general been delegated to the federal government by the Constitution. The decisions of the various state courts are sufficiently similar to make possible a discussion of "restraint of trade in common law" in all states together.

Contracts in restraint of trade

Since American common law was derived from English common law, American procedure stems from British legal practice. In some of the early interpretations in English Courts (beginning in the thirteenth century), all contracts in restraint of trade were regarded as invalid, but gradually this rule was relaxed. American common law came to distinguish between "unreasonable" restraints, which served no other purpose, and "reasonable" restraints, which were incidentally associated with practices in themselves acceptable and which did not work great injury to anyone. Thus when a peanut vendor sold his stand under a contract agreeing not to set up a rival business on the same corner within a year, the restraint on his freedom was an incidental, or, in the legal phrase *ancillary* restraint, reasonable under the circumstances of the agreement. This application of the rule of reason to ancillary restraints was clearly stated by Judge Christiancy in the case of *Hubbard v. Miller*:

If, considered with reference to the situation, business and objects of the parties, and in the light of all the surrounding circumstances with reference to which the contract was made, the restraint contracted for appears to have been for a just and honest purpose, for the protection of the legitimate interests of the party in whose favor it is imposed, reasonable as between them, and not specifically injurious to the public, the restraint will be held valid.

Two sets of interests are taken into account in this statement, those of the parties to the contract and those of the public. In most court cases focus tended to be on the interests of the parties to the contract rather than of the public. Nevertheless, contracts entered into directly for the purpose of raising prices were and still are generally held to be not only invalid (unenforceable in the courts), but evidence of criminal conspiracy against the public interest. Some unavoidable restraint of trade incidental to the pursuance of other purposes is, however, allowed, and such contracts can be enforced in the courts, when fair to both or all parties to the contract.

Integrations

One important reason for the development of the trust proper was the opportunity thus to evade the common law against contracts in restraint of trade, which invalidated and made criminal pooling agreements to control prices, restrict outputs, share markets, and in other ways to restrain trade. The new trusts flourished unchallenged for some years. In 1890, however, it was decided that the North River Sugar Refining Company had exceeded the rights granted in its charter by entering into such a trust agreement.² Two years later it was made clear that trusts aiming at monopoly were against public policy and hence the corporate

² *People vs. the North River Sugar Refining Co.*, 121 N. Y. 582.

charters of the member companies could be forfeited regardless of the specific powers granted in their charters.³

Since trust agreements were shown to be vulnerable under the common law an attempt was made to find a substitute for them by forming large corporations to take over the assets of the former competitors. These were holding companies formed in order to establish a monopolistic control in an industry. But in the case of the *People vs. the Chicago Gas Trust Company*⁴ decided in 1889 the court ruled that:

To create one corporation for the express purpose of enabling it to control all the corporations engaged in a certain kind of business, and particularly a business of a public character, is not only opposed to the public policy of the state, but is in contravention of the spirit, if not the letter, of the constitution.

This principle was reaffirmed in 1895 in proceedings against the Whisky Trust:⁵

There is no magic in a corporate organization which can purge the trust scheme of its illegality, and it remains as essentially opposed to the principles of sound public policy as when the trust was in existence. It was illegal before, and is illegal still, and for the same reasons.

Summary of the position of the common law

We may now summarize the usual position of the common law on these matters as it stood in the 'nineties and as, in essentials, it remains today:

1. Agreements aiming to raise prices are both invalid and criminal offenses.
2. Contracts entered into for other purposes, but incidentally restraining trade, are valid if the restraint is "reasonable."
3. Trust agreements entered into for monopolistic purposes justify forfeiture of corporate charters.
4. Corporations formed for the purpose of controlling other corporations and thus obtaining monopoly power are illegal as contrary to public policy.

The common law provided a large body of precedent that has since been followed in court decisions concerning collusive practices in restraint of trade; formal agreements in loose-knit federations engaging in monopolistic practices come under these decisions. On the subjects of large integrations and industrial mergers, however, there was little in the common law. The decisions against trusts and holding companies formed to monopolize an industry are not applicable to mergers and outright purchases of some companies by others.

³ *State vs. Standard Oil Company*, 49 Ohio, 1937.

⁴ 130 Ill., 268.

⁵ In the case of the *Distilling and Cattle Feeding Company vs. the People*, 156 Ill., 448.

Federal "Antitrust" Laws

The last quarter of the nineteenth century was a period of the growth of big "monsters" of corporate enterprise controlling economic, social, and political life to a degree unimagined before. Bolstered by tariff policies executed in their special interest, and stimulated by the development of railways that increased the size of the market, the barons of industry were developing large enterprises in many fields previously operated on a much smaller scale. Bigness was new, powerful, terrifying. Unscrupulous practices of powerful combinations such as the oil and sugar trusts had spread suspicion and hatred of these immense organizations. Practices in restraint of trade were growing in importance. Some of them were attacked in the state courts and held in check under the common law. There were, however, many difficulties in state action through either common or statute law. Laws and interpretations of these laws varied somewhat as between states. Public opinion was much stronger against combinations to restrain trade in some states than in others. Some states competed with others to get the fees derived from granting of charters to corporations, and hence there was an incentive to be lenient in the charter requirements and in prosecutions for practices of corporations that "restrained trade." Moreover, big corporations are powerful and to prosecute them is costly. Condemned in one state, a big trust could easily set up in another and perpetuate its monopolistic practices. Finally, state courts are limited in their jurisdiction to intrastate activities, and cannot do anything about interstate combinations and agreements in restraint of trade; the federal government on the other hand has jurisdiction where questions of interstate commerce are involved. For these reasons it was urged that some federal control should be introduced to support and supplement the activities of the states.

Provisions and interpretations of the Sherman Act

When in 1890 the congress passed the Sherman "Antitrust" Act as a frontal attack on big business monopoly and collusion, it clearly mirrored the popular demand for such legislation. The first objective of the statute was to enforce competition as the means of controlling industry, and to prohibit practices that interfered with these competitive controls. Since under the Constitution the federal government could do no more than suppress those practices that restrained interstate commerce, since it had no jurisdiction over intrastate commerce or production, the Act was set up to control practices in restraint of trade in interstate commerce. The two main provisions of the Act were directed against (1) restraints of trade, and (2) "monopoly" and "attempts to monopolize." They were stated as follows:

Sec. 1. Every contract, combination in the form of a trust or otherwise, or conspiracy, in restraint of commerce among the several states, or with foreign nations, is hereby declared to be illegal . . .

Sec. 2. Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of

the trade or commerce among the several states, shall be deemed guilty of a misdemeanor . . .

Like many other legislative declarations of policy, the Sherman Act has been subject to a long series of interpretations by the Supreme Court by which its effective significance in application is revealed. This process continues today.

The two chief questions in these cases are: (1) What activities are under federal jurisdiction under the interstate commerce power; and (2) how does the "rule of reason" apply to cases under the Sherman Act?

1. *Defining the scope of the interstate commerce power* is a complex problem. We may say roughly, however, that control of sales in interstate commerce has brought a business or a combination of businesses under the federal power, though until recently control of manufacture alone has not.

2. *The "rule of reason"* is an essential part of any application of a law that is as broad in scope as is the Sherman Act. In the early cases dealing with agreements in restraint of trade it was held that contracts entered into for other purposes, but incidentally restraining trade, were to be regarded as valid if the restraint was "reasonable"; all other contracts restraining trade were *per se* unreasonable. But this position was questioned in minority opinions at an early stage in interpretations of the law. It was argued by these minorities that the "rule of reason" should be applied to *all* restraints of trade, ancillary or otherwise. By 1911 this position was taken by the majority in the now famous *Standard Oil* and *American Tobacco* cases.⁶ Although the monopolies involved were broken up, the basis for the dissolution was not simply restraint of trade, but "unreasonable" restraint "aiming at monopoly." This broader interpretation of the applicability of the "rule of reason" had little effect on loose-knit agreements to fix prices, divide territories, control production, since these practices were now simply regarded as "unreasonable" restraints of trade (whatever the portion of the market covered).⁷ It is probable that the position taken in and since 1911 has led to more lenient treatment of integration into close-knit federations than would otherwise have occurred; but so long as the statute does not take either extreme view, on the one hand that no integration is permissible or on the other that all integrations are permissible short of complete (100 per cent) control of the industry, the Court with or without the rule of reason must draw the line somewhere between the two extremes. To "repeal the rule of reason" would not repeal the uncertainty.

⁶ *U. S. v. American Tobacco Co.*, 221 U. S. 106 (1911); *Standard Oil Co. v. U. S.*, 221 U. S. 1 (1911).

⁷ "No greater tolerance of the attempted justifications of such restraints after the adoption of the rule of reason was evidenced by the courts than during the period of its rejection. In the main and subject to minor exceptions, conduct which prior to 1911 was condemned as illegal restraint was held unreasonable *per se* thereafter and hence unlawful. The adoption of the rule of reason effected more of a change in theory than in actual practice." (TNEC Monograph No. 38, p. 8.)

Legislation of 1914

Dissatisfaction with uncertainties in the interpretation of the Sherman Act and mounting concern over the prevalence of "unfair" methods of competition led to the passage of two acts in 1914 directly intended to remedy these situations. The first of these was the so-called Clayton Act, the second was the Federal Trade Commission Act.

The purpose of the Clayton Act was primarily to clarify the Sherman Act in the light of the attitude taken by the courts under the "rule of reason." Toward this end, the Act enumerated a variety of specific offenses that were to be considered illegal and "unreasonable." These prohibited practices were: (a) price discriminations among different buyers of a commodity; (b) "tying contracts," which obligate the buyer of a commodity to buy certain other commodities or materials to use with it, such as typewriter ribbons with a typewriter; (c) acquisition by a corporation of more than a limited amount of stock in a competing corporation; and (d) interlocking directorates in directly competing corporations (except banks and common carriers) of more than \$1,000,000 capital, surplus, and undivided profits. But since each of these prohibitions was qualified by the phrase, "where the effect will be substantially to lessen competition or tend to create a monopoly," the Clayton Act effected very little change.

The Federal Trade Commission Act set up the body of the same name to provide a means of extending control over practices in restraint of trade and to carry on investigations to provide the basis for further legislation. The Commission's control was intended both to prevent the growth of monopoly through "unfair methods of competition" and to check exploitative practices of monopoly directly. The Commission was to hold hearings where illegal practices of either sort were suspected, either on its own initiative or on complaint of injured competitors, and to hand down "cease and desist orders" where the practices actually were found to be illegal. It was hoped that the Commission would both aid in checking monopoly practices and eliminate a large portion of the uncertainty arising out of the slowness of the courts in determining in particular cases what practices were "reasonable" and what "unreasonable."

From 1919, when the first major case involving an appeal from a FTC decision reached the Supreme Court, the powers of the Commission have been reduced to a very minor level so far as enforcement of the Sherman Act is concerned. The Supreme Court has consistently held that decisions of the FTC must be based on exactly the same considerations of "reasonableness" as those of the regular courts, and that it is for the courts, not the Commission, to determine what constitute acts that "unreasonably" restrain trade. Other factors have further limited the powers of the Commission. It cannot hold a practice illegal if only one competitor is injured, because it is authorized only to protect the "public interest." On the other hand, it cannot hold an act illegal that injures only the consuming public but does not injure competitors, because it is not author-

404 PUBLIC POLICY ATTACKING RESTRAINTS OF TRADE

ized to protect the consuming public. Thus in 1930 a company was required to desist from false advertising of obesity "cures," but the order of the Commission was overthrown by the Court because no other companies were selling similar cures and making similar claims.

The result has been that the Commission has been largely restricted to the hearing of a multitude of relatively minor cases on unfair trade practices, especially among small businesses. In this field, however, the Commission has done exceptionally valuable work. Its annual reports provide fascinating reading, revealing as they do the vast multitude of trade practices engaged in to mislead and defraud consumers and to gain business from competitors. Hundreds of orders have been handed down requiring sellers to cease advertising silk substitutes as real silk, machine-made lace as genuine Irish hand-made lace, patent medicines as sure cures for everything imaginable, retreaded tires as new tires, and other such practices. In recent years, the Commission has relied increasingly on "trade practice conferences" with industry representatives to eliminate such unfair practices; these joint conferences draw up programs of fair competition that are then adopted as the standard for the industry. Such voluntary conferences have proved much more effective in covering a wide range of cases than the older practice of investigating single cases or waiting for complaints from competitors. The trade practice standards established at such conferences, although they are primarily intended as protection for members of the industry against each other, in effect provide a considerable protection for the consumer.

Enforcement of the Antitrust Laws

There has never been a full-fledged effort to enforce the antitrust laws. Until recently, staff has been inadequate for investigation of offenses, and enforcement has therefore been erratic and infrequent; erratic and infrequent prosecutions have exaggerated uncertainties among businessmen concerning the content of the law in many situations; and penalties have frequently been either too light or too severe.

The new attack on restraints of trade since 1938: staff and scope

The inadequacy of staff and resources available for antitrust enforcement has been a glaring defect of public policy. Even the famous "trust-busting" campaign under President Theodore Roosevelt at the turn of the century was conducted with a staff of only five lawyers and four stenographers. The results of that campaign were dramatic, focusing on the most gross contraventions of the law by big business; but such an attack cannot rid the economy of the many restrictive practices with which it is riddled in hundreds and thousands of small corners, practices that add up to a tremendous total effect. Although the staff of the division was gradually expanded, it was impossible until very recently to conduct more than four antitrust cases at one time.⁸ Interest in enforcement waned,

⁸ Edwards, Corwin D., *The Function of the Antitrust Laws in the American Economy*, Address before the Trade and Commerce Bar Association, February 29, 1940. Mimeographed, p. 7.

and initiative in bringing up cases was left largely to private complaints.

At the end of the 'thirties Thurman Arnold became head of the Anti-trust Division and the government undertook a major campaign to bring life into the antitrust laws. The department was expanded until by 1941 it had a personnel of more than three hundred. With this expanded personnel it was now possible to change the methods of approach in two very significant ways:

1. Instead of waiting for complaints, the Antitrust Division took the initiative in ferreting out violations of the law in many industries and many places.

2. A frontal attack on restrictive practices in all the branches of an industry replaced the case by case approach of earlier days. The prosecutors included restraints practiced by labor⁹ and agriculture as well as by business groups. Especial drives were first directed against restraints affecting housing construction and against restraints in the processing and marketing of foods. Almost unbelievable collusive arrangements were uncovered in the investigated industries.

The breadth of the investigation created an impact that would have been impossible on the case by case approach of earlier times. Persons who had long argued that the antitrust laws were unenforceable were awakened with a shock; businessmen, farm producers and marketers, and labor union leaders, who had become accustomed to public tolerance of monopolistic practices, were even more rudely awakened; but the general public, learning for the first time of the collusion, monopolistic restraints, and graft and corruption so prevalent in the investigated industries, possibly received the biggest shock of all. The broad enforcement of the law in each industry probably began to have some preventive effect, since offenders no longer considered themselves as relatively unlikely to suffer prosecution for illegal acts; and the Division received considerable cooperation from businessmen because the more systematic coverage was clearly both more equitable and more constructive in its effects.¹⁰

⁹ Application of the antitrust laws to activities of labor groups has been sharply limited in recent Court decisions. See Chapter 33.

¹⁰ Mr. Edwards' comments on the importance of this simultaneous attack on restraints throughout an industry are important enough to deserve extensive repetition: "Isolated anti-trust cases may end particular restraints, at least for a time, but they accomplish little else. If in the housing drive we had prosecuted the Long Island Sand and Gravel Association this year, the Detroit tile contractors next year, and the Chicago stone cutters' union the year after, we would have done substantially nothing to reduce the cost of building and would have exercised no real deterrent effect upon the various groups we have not yet prosecuted. By summoning eleven grand juries simultaneously in key cities through the country and by presenting to them the related restraints of manufacturers, distributors, contractors, sub-contractors, and labor, we have undertaken prosecutions broad enough to accomplish something. We have begun to reduce housing costs. In Pittsburgh the second unit of a low-cost housing project, for which bids were let after our work began, is to cost \$148 per room less than the first unit, which was begun prior to

But even at its peak, just before and in the early stages of the Second World War, the Antitrust Division was still inadequately staffed to really do the job that lay before it. A policing of the ground gained requires the provision of permanent field staff stationed in the various states to keep an eye on developments. Such a staff could do much to prevent a recurrence of illegal practices once they have been cleaned up, and to prevent the development of new budding restraints into hardy and tough weeds that would be difficult to eradicate. This was not yet possible when the war came and put something of a damper on the growing effectiveness of the Antitrust Division.

The relation between enforcement and uncertainty

One of the most common complaints against the antitrust laws is that their meaning is too uncertain so that businessmen and other groups do not know where they stand. Insofar as this is a valid complaint it arises for three major reasons.

First, because the act is very broad, there must always be some marginal cases that are doubtful; this is inevitable in the formulation of a program with sufficient flexibility to meet the continuously changing conditions, and the variety of techniques of restraint of trade that develop in a dynamic economy.¹¹ While some specifications may be useful, an attempt to specify in all particulars just what practices would be illegal would at once open the door to evasions and would prohibit, unintentionally, practices of definite social usefulness.

Second, an impression of uncertainty in application of the law arises when its enforcement is erratic. One businessman is prosecuted while others engaged in the same or similar activities get by unscathed. The frontal attack on an entire industry is a means of eliminating this type of uncertainty and arbitrariness.

Third, when resources for enforcement are inadequate prosecutors must necessarily focus attention on the most obvious and flagrant cases, so that many issues are never cleared up by being brought out for an airing.

our investigation. We have induced many building groups to clean house voluntarily, not only in the cities in which we are at work but elsewhere. We are creating an environment in which particular industries and trades can proceed on a competitive basis without feeling that they are merely being suckers for some other group's benefit. Because this kind of program is constructive, most of the groups in the building industry have given it warm support without regard to any risks that they may personally run, whereas if we were engaged in haphazard prosecutions we would encounter last-ditch resistance. The constructive aspects of the policy and its persuasive influence upon the opinion and behavior of the industry are due to the fact that we have been able to prosecute upon something approaching an adequate scale." (*Ibid.*, pp. 7-8.)

¹¹ In speaking of this problem to a group of lawyers, Mr. Edwards remarked: "It is not feasible to write into a statute the details of such a law. An effort to do so probably would mean that business men, well-advised by the members of your ingenious profession, would find little difficulty in evading the law's purpose and that many innocent activities would be forbidden by inadvertence." (*Ibid.*, p. 10.)

Penalties under the Sherman Act

The government may follow two kinds of procedures in prosecutions under the Sherman Act. The first is a procedure in "equity" under the civil law. Such a procedure culminates merely in an order to cease and desist from a continuation of the illegal practices. There may be dissolution of an organization with reorganization of the industry, but there is no real penalty here to make the offenders worse off than had they obeyed the law right along. Such action exercises no deterrent effect on other businessmen who may be engaging in the same kinds of activities. The second is a criminal procedure. Prosecution as criminals, with the possibility not only of fines but of imprisonment, with all the stigma attached to such an experience, may be a very effective penalty. But in some cases it is regarded as a penalty that is too harsh, and it may therefore prejudice a court against declaring an offender guilty. Whether penalties in civil law can be and will be developed which are adequate to handle the situation remains to be seen; revisions of the law to provide such penalties as alternatives to criminal prosecutions have been recommended by Mr. Arnold and other members of the Antitrust Division.

Can an Antitrust Program Preserve Competition?

Attitudes toward the future of American economic life vary from enthusiastic optimism to deep gloom. Attempts to examine and formulate criteria of social welfare, and to study their implications for policy making are permeating wide circles where for many decades these matters have received relatively little thought. Extended surveys of the structure and operation of the American economy provide a vast body of information previously unavailable. Thousands of black marks have been scored up against that economy, and hundreds of specialists are giving concentrated attention to erasing these blots on the scutcheon of Democracy. The problems are complex. The goals of encouraging abundant production and progress seem simple enough in themselves, but the attempt to realize them presents vast problems. Their solution requires the development of techniques for breaking down restrictive practices while retaining the economies of large-scale production. Their solution requires a simultaneous solving of the problems of transitional adjustment and protection for those who become stranded in industries and productive activities that are on the wane in the dynamic evolution of economic life. Out of the maze of material that we have discussed in the immediately preceding chapters, what guiding threads may be drawn to lead us to wise solutions to these problems?

Competition is a process by which in a free enterprise economy conflicts between different groups are resolved and the whole economy brought into balance; it is the mechanism that keeps open the channels of trade, thus stimulating improvements and abundant production. The antitrust policy that has been developed in America is designed to preserve this

competitive mechanism as a guarantee of protection to the population as a whole. Opinions vary as to the degree to which competitive practices can be preserved. We are now in a position to evaluate the role of antitrust laws in preserving competition.

The meaning of "competition" in the application of the antitrust laws

"Competition" as the term is understood by the ordinary businessman is a much broader term than the "pure competition" of economic theory. No sane person would today argue that it is feasible to create an economy characterized in large part by "pure competition," even though in some phases of economic activity conditions approximating pure competition are clearly evident. But many people argue that it is feasible to preserve and to increase effective competition as a regulator of economic adjustments. What is this competition that we are seeking to preserve?

To many people the term "competition" means simply that two or more sellers are trying to attract the same customers or that two or more buyers are bidding for the same goods or services, *and that these rivals are active in their rivalry and have not entered into unified programs of collaboration.* Frequently there is the additional implication in considering competition that there is *freedom for new firms to enter into a market*, and when people are arguing for preservation of free competition in this respect they usually mean also that these new firms should be allowed to appear and grow without having to encounter and overcome unfair discriminatory practices directed against them. While these concepts are very hazy in common speech and thought, they nevertheless approximate closely certain of the analytical concepts that were introduced in earlier chapters of this book.

1. The term "aggressive competition" was used to describe competitive behavior that would follow if each of a group of rival firms simply reacted to existing market conditions without considering the effects of its actions on the policies of rivals. We saw that in oligopoly and oligopsony situations price competition (and frequently quality competition as well) tended to be nonaggressive even when there were no formal collusive agreements, and that the results of such tactics were very similar to those that would follow from more formal collaboration between rivals. In the common use of the term the nonaggressive policies involved in tacit or formal collaboration would be regarded as noncompetitive; and the "preservation of competition" would then involve the enforcement of conditions that would generate "aggressive" competition as we have defined it. (See Chapter 23.)

2. Popular and scientific use of the phrase "freedom of entry" are very much the same. Such freedom exists when a new firm is free to enter an industry or a production area on the same terms as those already established in that market. Interferences with freedom of entry take all the many forms described in preceding chapters, and especially in the

discussions of foundations of power in Chapter 26. A policy of preserving competition would then involve the elimination of all "unreasonable" restrictions on the entry of new firms.

Oligopoly and restraint of trade in the antitrust program

The pursuit of uniform and coordinated programs in pricing and market sharing tends to some extent to characterize most industries in which the number of firms is small. When there is overt collaboration, antitrust enforcement is relatively easy, but when there is merely tacit following of a leader or consideration of the reactions of rivals the problem is much more difficult. Let us consider these two kinds of situations in some further detail.

Out and out agreements between different firms to pursue coordinated market policies are often considered to be old-fashioned and quite out of the picture "now-a-days," but to harbor this idea is to harbor an illusion. In fact, as some of our previous discussions have indicated, this is a very common practice at the present time. Members of the Antitrust Division claim that such cases can easily be tracked down if the staff of the Division is adequate, that agreements of this kind, however secret they may be, leave inevitable traces.

If there are many persons in the conspiracy, one of them is almost sure to develop a grievance and turn state's evidence. If the plan is complicated, it is almost certain to leave written records. If no letters or memoranda are written, there is evidence that meetings were held and evidence of the identity of action which followed the meetings. If files have been stripped, stray carbons and references to the missing documents in other documents make it possible to trace what happened. Once proved, a collusive agreement is relatively easy to terminate by law, since it operates through a system of joint action which may be stopped.¹²

Once such a case is tracked down there is little doubt as to how it will be handled in the courts. Collusive agreements to raise prices, curtail outputs, and share markets have been consistently regarded as unreasonable restraints of trade. There is no possible defense that such practices involve important economies of production, as is argued in cases of big business consolidations.

Insofar as uniform policies are due merely to following a leader or to consideration of rival reactions by each of a small group of firms, results may be quite similar to the clear cases of collusion, but it is extremely difficult to distinguish between actions that are legitimate and actions that interfere with competition. In these cases there is rarely a sufficiently clear basis for the bringing of an antitrust charge. Only when there are overt threats and business practices clearly designed to destroy firms that fall out of line is there much that can be done about the situation. Yet the individual pursuit of nonaggressive practices in the markets may lead to high selling prices (or low buying prices), restricted

¹² Edwards, Corwin D., "Can the Antitrust Laws Preserve Competition?" *American Economic Review*, March Supplement, 1940, p. 175.

production, and high profits that persist over very extended periods of time. To break up such tacit collaboration would be impossible under the existing interpretations of the antitrust laws, and it is probably not feasible to develop any kind of law that would get at such practices directly.

The problem of size

The most basic issue in the attempt to preserve competition centers in the problem of size of enterprise. Is it feasible to preserve efficient production in business units small enough so that most markets can conceivably be characterized by active competition? Opinion is unanimous that some industries, as local light and power companies, must be regulated monopolies; we have already a long tradition of government regulation in "public utilities." The spheres in which the issue of efficient size as affecting market controls looms large are those in which there tend to be a small number of firms. It is probably impossible to force aggressive competition where firms are few. We can cope with tacit oligopoly restraints on competition only by removing their cause, the existence of a small number of rival firms.

Action directed against the formation of close-knit federations is essentially preventive action; it is intended to prevent the formation of business units that will make possible domination of a market and consequent restrictive price and production policies. Public policy in this matter has, however, been wholly inadequate. It has been confined in large part to cases in which the business integration would involve consolidation in one firm of a very large share of the total productive capacity in an industry, but has paid little heed to the implications of increasing size where there were still several major rival firms. It has focused on particular kinds of integration, the trusts and holding companies, but has failed to curb in any way the development of big business consolidations through mergers and acquisitions. The tremendous splurge of such integrations in the 'twenties greatly increased the difficulty of enforcing competitive practices in American industry by facilitating effective tacit collusion among small numbers of large concerns. Those objecting to proposals to curb mergers argue that such integrations make possible much greater production efficiency and hence lower costs; but there is very general evidence to show that most mergers (as well as holding companies) are profitable primarily because of the monopolistic controls resulting rather than because of any lowering of production costs. Whether there will be any real attack on this problem in the years following the war remains to be seen; but this is clearly today a major gap in a program designed to preserve competition as a mechanism for controlling economic adjustments.¹⁸ Finally, how far a well-enforced pro-

¹⁸ Mr. Edwards of the Antitrust Division expresses himself on this matter in very strong terms:

"A solution of the problem of size is the heart of the question whether competition can be preserved by the anti-trust laws. Without competitors there can be no

gram preventing mergers can go in preventing concentration of production in a few enterprises in an industry will vary from one industry to another. Whether such a policy, supported by other aspects of a program designed to preserve competition, can free the major part of the American economy from tacit oligopoly collusion remains an open question.

Some people believe that economic changes are rapid enough so that by preventing the further development of large business concentrations in the future we shall be able to handle the problem of size as a basis of market controls. Other writers suggest going further than this by breaking-up existing large enterprises into smaller units. Such a policy could be carried out in some industries without any significant loss of production efficiency, since the principal economies of size are the technological economies embodied in the operation of separate plants or establishments, and big business organizations are commonly firms including a number of separate establishments. But in many sectors of the economy this is not the case. The practical difficulties, especially the political difficulties, involved in carrying out such a sweeping program are terrific, though it is debatable whether or not they are overwhelming.

The problem of bottlenecks

Bottlenecks in economic activity create and support monopolistic conditions that are then difficult to handle under any program designed to foster active competition. In Chapter 26 we listed five types of bottlenecks: transportation facilities, marketing channels in buying and selling, raw material and power sources, patents, and labor supply. Closely related are the exclusive positions created by some of the laws setting minimum standards—for example, certain features of local building codes. The remedy for these under the antitrust laws is slight indeed. Much of the centralization of control of transportation and of marketing channels reflects genuine economies of large enterprise; insofar as this is true the problem of size emerges again in a form that suggests that some policy of regulation instead of enforced competition may be required.

competition; and with the present trend of corporate growth the disappearance of most competitors in many industries can be foreseen. Expansion for efficiency's sake probably is a self-limiting process; but expansion by fusion obviously is not. If this trend can be halted or reversed, the other problems involved in the enforcement of the anti-trust laws will amount to no more than the ordinary problems of making an administrative agency effective. If, however, the scale of business is to continue to grow not only by expansion but by unchecked merger, it will be necessary to accept oligopolistic price policies as typical and to fight the battles against complete monopoly only as last-ditch struggles in a war already lost. Such a prospect would require the community to consider adopting types of industrial control which operate directly upon the policies of corporations, in recognition of the fact that a giant concern with powers not limited by its rivals in the market is a quasi-public enterprise. Such a development would involve, not the making of exceptions to the antitrust laws, but the preservation of the areas of competition as exceptions to the general economic policy. I do not relish that prospect, for it would require a vast machinery of control over the detail of industrial decisions; and in my judgment governmental control is one of our social resources which should be carefully economized." (*Ibid.*, p. 179.)

Similar problems arise in connection with controls over raw materials and power sources; though antitrust action can by itself do something. In the case of the Aluminum Company's monopoly of bauxite two remedies outside the antitrust program suggest themselves, both consistent with a program designed to further competition: repeal of the tariff that has prevented foreign bauxite from competing in the markets of the United States, and encouragement to the financing of producers using a new process that makes possible aluminum manufacture out of other clays. The importance of patents as bases of monopolistic power has received increasing attention from students and the general public; here antitrust policy might be bolstered by a revised patent law. Public policy toward the manipulation of labor supply as a bottleneck must be part of a whole broad program dealing with labor problems.

Clearly the breaking of bottlenecks cannot be accomplished by the Antitrust Division alone. It can accomplish only a small part of the task. Some bottlenecks may be broken through other aspects of public policy designed to foster competition. Some are unbreakable, and call for direct government regulation which will make the use of the facilities (for example, oil pipe lines) available to all on equal terms; in this way the inevitable centralization of control at one point will be prevented from spreading into a huge web of monopolistic power. Each case requires examination in detail, though to all we may bring the same broad criteria of fostering abundance in the full use of resources.

The income problem

In an attempt to evaluate the possibilities of preserving and encouraging competition through antitrust policy, we cannot ignore the awkward problem of income adjustments for those who get caught in stranded economic positions. It is this focus on the income problem that explains in large part the recent government-fostered cartelization in agriculture and coal production that has been described in previous chapters.¹⁴ How can a policy be formulated that will facilitate the shifting of human resources out of these spheres into productive use, and at the same time prevent severe suffering for those most immediately involved? Can such a policy operate consistently with a program such as antitrust, designed to foster active competition?

Public policy may be directed toward removing the cause, and curing the malady. The causes that bring about problems of economically stranded groups are partly the broad changes that occur in a dynamic society over long time periods, partly the changes associated with wars, partly those associated with general business fluctuations (commonly called business cycles). They are also due to monopolistic practices that lead to the building up of excess capacity in an industry. Adjustment to gradual long period changes are easily brought about by a consistent program to foster mobility out of contracting and into expanding eco-

¹⁴ Chapters 26 and 27.

conomic areas—largely through providing increased information, vocational training and vocational guidance. These programs are in fact ways in which government can help bring about economic conditions that more nearly fulfill the requirements of a competitively operated economy. The stranded groups resulting from war activity are not so easily prevented, and are in a sense publicly or collectively caused. Here there falls on government an especial responsibility for providing re-training for the stranded people and thus creating the mobility that is necessary for readjustment. This again is consistent with the bolstering of an antitrust program designed to foster competition. The problems created by business cycle changes will be discussed in later chapters;¹⁵ but here we may merely note that removal of monopolistic obstructions is one of the most important points of attack generally argued to prevent or modify general business fluctuations. Insofar as stranding is due to monopolistic practices, antitrust policy obviously contributes to preventing the occurrence of this situation.

Unfortunately, these suggestions are not or at least have not been regarded as enough; nor have they been effectively tried out. It has seemed easier simply to step in and help the depressed industry get a higher income by collaborative price and output policies. This policy has been and is being tried, as in the case of bituminous coal. It has usually been justified at first as a temporary transitional program; but experience seems to indicate two things. Like the tariff, once initiated it is difficult to get rid of; and instead of encouraging mobility out of the contracting market it may actually exaggerate the surplus problem and therefore the need for assistance. New justifications of the intervention appear—principally the argument that through collaboration greater stability can be attained to the interest of producers and consuming public alike. It is also argued that where many small producers are dealing with oligopoly or oligopsony groups, they need the protection of collaborative action of a cartel type. These arguments simply lead us back to the problems of size and of bottlenecks discussed in the last two sections. They are not a part of the income problem as such.

Antitrust in a changing scene

The preservation and encouragement of active competition is only one way of seeking abundance and progress; though many would regard it as the only way which is both feasible and compatible with the functioning of democratic institutions as we know them in our individualistic inheritance. Antitrust is the main but not the only cog in a mechanism for implementing competitive controls.

As the structure of economic life becomes increasingly complex, and as control of economic actions shifts to larger groups and more powerful units, the task of an Antitrust Division becomes increasingly difficult. At the same time we see more clearly than ever before the necessity of

¹⁵ Chapters 44 and 45.

integrating antitrust policy with other aspects of public policy. There are many conflicts in public policy today—a few of them have been indicated in the last few chapters, but there is much that lies beyond our brief comments. Antitrust policy is applied in an economic setting the structure of which is changing; it is applied as one aspect of public policies that are ramifying into more and more detailed governmental control of industry. In some economic spheres public policy may be coordinated along many fronts to support the effort to maintain and encourage competition. In others we are forced to seek abundance and progress in other ways.

The degree of success attained in an antitrust program will depend not only on that program itself, but on how actively it is enforced and how well other aspects of public policy, such as the handling of patents, are coordinated with the antitrust program. To the extent that it succeeds in its task, antitrust action clears the way for efficient use of resources through individual initiative with a minimum of governmental control and pressure politics.

CHAPTER 29

Public Utility Regulation

The cases just cited show that, while, ordinarily, free competition in the common callings has been encouraged, the public welfare may at other times demand that monopolies be created. Upon this principle is based our whole modern practice of public utility regulation. It is no objection to the validity of the statute here assailed that it fosters monopoly. That, indeed, is its design.

These are the words of Mr. Justice Brandeis in his famous dissenting opinion in the *Oklahoma Ice Case*. The principle expressed grows out of the characteristics of most public utilities such as water, electric, and telephone companies. They are believed to be in some special manner or degree "affected with a public interest." They usually share with some of the industries regulated only under the general antitrust laws certain traits that make pure competition impossible, and they are usually more efficiently operated as simple monopolies than as duopoly or oligopoly industries.

The distinctions between industries that are and those that are not public utilities have not always been consistent and no clear separation can be made in our analysis. Some businesses have been considered in a special class as "public callings from time immemorial" in the legal phrase. Others have been added to the list in more recent legal history. Some industries may be subjected to special regulation, as of prices and rates, and yet not be made monopolies by the granting of public utility franchises. These are in an intermediate zone between the industries subject only to the general antitrust laws, and those industries that are generally classed as "public utilities." The final diagnosis as to classification of an industry is the opinion of the court. This opinion is based on precedent, on the economic characteristics of the industry in question, and on the political-social attitudes of the currently dominant policy makers.

Those economic characteristics of an industry that are most likely to result in the treatment of the industry as a public utility were suggested in the first paragraph of this chapter: (1) The optimum scale of enterprise is so large relative to the size of the market that pure competition is impossible. Frequently in fact simple monopoly is more economical than duopoly or oligopoly; that is, this is a decreasing cost industry. (2) The services of the industry may be considered "essential," this

essentiality being partly in the nature of the service provided, partly in the absence of any satisfactory substitutes. Thus no one food is usually regarded as "essential," but water is. Milk has been a subject of considerable dispute. When consumers have no recourse to close substitutes for "essential" goods, they are easily exploited. (3) Duplication of services may lead to inefficiency and inconvenience to consumers. As those who have seen a flank of four streetcars on the streets of San Francisco are aware, two streetcar lines congest traffic unduly. Two sets of power lines blacken the skyline of Seattle as well as increase the costs of providing electricity. Residents of Mexico City and even of some small towns in the United States are forced to have two telephones if they wish to be in touch with friends subscribing some to one system, some to another.

The list of industries generally considered public utilities has slowly expanded with changing social attitudes and rising standards of living. Water, gas, electricity, street railways and buses, and telephone and telegraph service are the industries today generally accepted as public utilities properly subjected to stringent control. In an intermediate category are the railroads and other "common carriers" such as interstate truck and bus lines. Air transportation appears gradually to be falling into the same group. In these intermediate cases there appears to many observers to be a place for some competition among a few firms at the same time that it is generally agreed that a considerable amount of governmental regulation of rates and competitive practices must be exercised; the conflict arises over how extensive this regulation should be. Still more indefinite is the case of milk distribution in urban communities, an industry in which competition has dwindled steadily and in which the wastes of cost-raising rather than price-reducing competition are obvious and important. There is reason to suppose that this may be the next industry generally accepted as at least a quasi-public utility, although the speed with which such changes occur varies widely from one case and period to another.

The relative magnitude of the public utilities in our economy today is often underestimated. There are well over 20,000,000 telephones in use in the United States, with millions of miles of wires. Over 80 per cent of the homes in this country are equipped with electric service. Gas and central water supply service are available in virtually all urban centers and in many rural communities. Rail, truck, bus, and telegraphic facilities reach into every corner of the nation. In total, the public utilities have recently accounted for around 12 per cent of the income produced in the United States, an amount approaching \$10,000,000,000 in 1941. Measured in terms of employment their relative importance is smaller, because of the small proportion of labor to material agents in the utilities; for the same reason, in terms of physical capital employed the utilities bulk correspondingly larger. Yet the railroads alone employ directly over 1,000,000 persons.

The public utilities account for a large and vital portion of our economy under complete or very substantial government control. Their impor-

tance emphasizes the significance of the regulatory problems involved. Tens of millions of consumers, billions of dollars of investment, millions of public utility employees are intimately affected by the policies adopted. With the apparent trend toward classification of more industries as utilities, present policies take on an additional importance since policies once adopted are later cited as precedents for further actions.

Regulation of public utilities has many facets. Important among them are the following: (1) grants of special privileges supporting or strengthening monopoly position; (2) protection of consumers by control of quantity and quality of service; (3) protection of consumers by control of rates charged, and (4) protection of consumers and the general investing public from financial manipulations by "insiders" in public utility holding companies. Before examining these problems, however, we turn briefly to a consideration of the machinery of public utility regulation.

The Machinery of Regulation

The basic public utilities have long been subject to special governmental regulation as being "specially affected with the public interest." Under the common law they were viewed as industries that, generally having an exclusive position in the field with the right of eminent domain, were properly required, within the limits of their capacities, to serve all comers impartially at reasonable prices. Early in the history of the United States, many municipalities and a few states attempted to regularize and further establish the rights and duties of public utilities by enacting statutes and ordinances for this purpose and by granting franchises bearing both special privileges and duties. Such control methods seldom proved satisfactory for long. Often the provisions were carelessly and loosely drawn; even when they were not, their inflexibility in periods of rapid change led to violation or simply disregard of their terms by all parties.

Gradually the need for more flexible forms of regulation gained wide recognition, and through the middle 1800's several states established regulatory commissions of one sort or another. These early commissions were solely advisory in nature, their chief functions being to collect and disseminate information and to make recommendations to state legislatures. Most of them were originally railroad commissions, and following the Civil War a gradual trend toward granting some mandatory powers to the commissions was apparent.

Despite their limited powers and scope of action, the work and experience of these early commissions provided a valuable foundation for the wider powers that they have since accumulated. From the start of the century commission powers began to be extended beyond the railroad field into the other utilities, primarily water, electricity, telephone, and gas. Beginning with the Public Utility Laws of Wisconsin and New York in 1907, commission regulation began a rapid growth that has culminated in almost complete control through this medium in the entire United

States. Superficially, the form of control through municipal ordinances and franchises continues in some areas, but with questions of rate and service controls largely removed to the state commissions, municipal control is of limited significance. State utility commissions have steadily been strengthened, and in recent years this increase in power has been accompanied by a serious effort to get more competent, impartial persons on the commissions to remedy a weakness in personnel that has long been widely recognized in the utility commissions of many states. At present, in the more advanced states commissions have virtually complete power over the utilities that they regulate, as to granting and withdrawing of charters, as to type of service provided, and as to rates charged. In other states, commissions' powers are substantially less far-reaching.

Superimposed on this emergence of the state commission form of control has been increasing regulation by the federal government. The Interstate Commerce Act of 1887 established federal control over interstate carriers through the commission form, but it was not until the decade of the 1930's that the federal government finally acted to take control over the prices of electricity and gas sold at wholesale in interstate commerce and to encourage a more integrated and rational public utilities system for the nation. The federal control is administered in large part by the Federal Power Commission, which in addition to its own duties gives frequent aid to the various state commissions. From a somewhat different angle the Securities and Exchange Commission exercises a considerable degree of federal power over the utilities through its activities in regulating the character of corporate securities sold on the major exchanges and through its specific power to investigate and force public utilities holding companies to simplify and integrate their structures.

Privileges Granted Public Utility Companies

Two important privileges are frequently granted public utility companies: (1) the exclusive right to operate in certain markets and (2) the right of eminent domain. Both of these kinds of grants may benefit the general public under certain conditions and when accompanied by careful regulation of the activities of the firms involved.

The exclusive right to operate

It may at first seem strange that entry to an industry should be confined by law to just one firm. This may, however, be clearly in the public interest under both or either of the following conditions: (1) in a "decreasing cost industry" and (2) in an industry in which consumer interests in quality and convenience of service are much better served by one firm only than by two or more firms.

1. When the optimum scale of enterprise is very large relative to the size of the market, pure competition is obviously impossible. In most public utility industries, moreover, production by one firm is more eco-

nomical than either duopoly or oligopoly; the industry is characterized by "decreasing costs." Yet if there were nothing prohibiting the entry of a second firm it might quite readily appear, sharing the market with the first enterprise. This would in no way benefit consumers since competition between two firms offers no safeguard, and if the industry were truly a "decreasing cost" industry any given output of the two firms together would be produced at greater cost than had one firm produced it alone.

2. At least equally important is the inefficiency and inconvenience to consumers that may result from duplication of some types of services. Examples of the inconveniences of such duplication have already been cited. Provisions granting exclusive rights to a single company are obviously a method of preventing these developments.

The right of eminent domain

There are many circumstances in which the interests of the general public require that some private property owners be forced to give up their property in return for a compensation at a "fair value." To lay out a railroad, for example, would be an almost insurmountable task if occasional stubborn property holders dotted along the line refused to give up their land for any reasonable sum (or even for any sum at all). The right of eminent domain is the right to compel such an individual to sell out where it is "in the public interest" that he do so. Many public utility companies may exercise this right under certain important limiting conditions designed to prevent unjustified exploitation of their power.

Regulation of Quality and Quantity of Sales

A number of very complex considerations are involved in the attempt to protect consumers of the products of public utilities. Among them are questions as to the qualities of services provided, and where, to whom, and in what quantities they shall be made available. Controls over quality of service are very weak and largely indirect by-products of rate regulation. Controls over the scope of sales are both direct and indirect.

The first protection to consumers necessary in the granting of exclusive privileges is the guarantee that all consumers who wish to pay for the service shall be granted the privilege of using it without unfair discrimination. The utility is thus required to serve all consumers wishing the service at the established rate. This requirement is so obviously important as to require no further comment.

A major problem sometimes arises with regard to the establishment of new branches of a utility in previously unserved areas. Rural electrification is a striking example. There has been no clearly established and generally accepted criterion on which to base decisions of this sort. Increasing heed is, however, being paid the thesis that where addition of a new branch of utility services can add more to revenue than it will add

to costs the expansion "should" be undertaken; this would be consistent with the "best" allocation of resources even though the new group served would pay less than its "share" of the general overhead costs. In actual practice understanding has been incomplete and policy has been mixed. In some cases commissions have taken steps to require utilities to extend their services into new areas, even though the additional costs could not be covered from the new consumers, and were assessed back on other parts of the companies' activities. Likewise, a utility may not suspend any part of its services without permission of the commission, and in a number of cases such permission has been refused.

Indirectly quantity of services offered is controlled through the regulation of rates charged for utility services; these rates, once determined, become incidentally a basis of control of quantity of service because the utilities are required to offer all that is wanted at the established rate.

Rate Regulation

Public utility "rates" are the prices charged for utility services. Of what have these rates been made? Probably of a large portion each of power pressures and ignorance with a small flavoring of analytically formulated economic criteria. Unfortunately it is only recently in most states that commissions have made real strides toward high standards of competency and social responsibility in membership and operating personnel.

The basis of all utility rate regulation for years has been the famous case of *Smyth vs. Ames* (1898) in which the Supreme Court decreed that rates must be set so as to provide a "fair return on a fair value" of the public utility investment. Any lower return would be confiscatory under the Constitution. Reference to the promotion of consumer welfare is conspicuously absent.

This pronouncement of the Supreme Court has left two basic problems to be determined: what is a "fair return," and what is a "fair value" on which the return is to be calculated?

The problem of a "fair return"

The question of a "fair return" has been easily resolved by the courts and applied by the commissions. For the period of the 1920's the various state courts generally held 7 per cent or 8 per cent to be a fair rate of return to investors. Through the bitter depression of the early 'thirties this position showed little change, and it was not until the latter half of that decade that rates of around 6 per cent came to be considered proper by the courts. Through prosperous periods the returns dictated by the courts on investments in utilities have been, to say the least, very generous, in view of the virtually riskless nature of such investment. In depression the courts' "fair" rate of return on utility investment has been nothing short of phenomenal as compared with other investment opportunities of comparable risk. Yet the courts give little indication of

moving more rapidly now than in the past in adjusting "fair return" figures for public utilities to changing market returns.

The problem of a "fair value" of investment

With the entire problem of utility rates generally conceived in terms of a "fair return on a fair value of investment" and with the "fair rate of return" essentially held constant by the courts, the determination of a "fair value" of utility properties has for decades held the center of the stage for regulatory commissions and private public utility experts alike.

In the same case of *Smyth vs. Ames* that gave the "fair return on a fair value" formula, the court enunciated the following distressingly ambiguous procedure for determining "fair value":

. . . in order to ascertain that value, the original cost of construction, the amount expended in permanent investments, the amount and market value of its bonds and stocks, the present as compared with the original cost of construction, the probable earning capacity of the property under particular rates prescribed by statute, and the sum required to meet operating expenses, are all matters for consideration, and are to be given such weight as may be just and right in each case. We do not say that there may not be other matters to be regarded in estimating the value of the property.¹

Beyond this listing of considerations the court has seldom been willing to go, insisting that the value of property is a matter to be determined by "informed judgment." In the last analysis, the "informed judgment" of the court supersedes that of the commission, whose members presumably are persons closely acquainted with the utilities field and who spend a majority of their time investigating precisely these problems. That rate setting and the determination of "fair value" have shown a remarkable lack of economic analysis and understanding is not surprising under these circumstances.

It would serve no purpose here to present in detail the obtuse and legalistic controversies concerning the determination of fair value for rate-making purposes. All of the considerations mentioned in *Smyth vs. Ames* have been used from time to time, but attention has been primarily centered on two criteria—the original cost of production of the property, and its present reproduction cost. The property value figure cannot be obtained by capitalizing the earnings of the firm, since the earnings themselves are in considerable part dependent on the rates that the commission is trying to set. The par value of the corporate securities is an obviously unsatisfactory measure in view of the large amount of overcapitalization and stock-watering in the field, while the market value of the securities is open to the criticism noted above, that their value depends on exactly the rates being set plus the vagaries of market price fluctuations. This leaves in practice only the original cost and reproduction cost of the property of any real usefulness. And neither of these is satisfactory, each having serious drawbacks.

The ascertaining of either "original costs" or "reproduction costs" is

¹ 169 U. S. 466, 546-547 (1898).

an extremely complex practical problem; but "original costs" involve fewer difficulties than do "reproduction costs." Once original costs have been estimated it is only necessary to make adjustments year by year to take into account depreciation and replacement; whereas reproduction costs would involve complete new estimates with every change in prices and character of the equipment to be used.

Precisely what is meant by "original costs"? Frequently the commissions have simply sought to determine what the recorded costs of producing the existing plant and equipment are, but such costs were frequently wastefully high and padded with promoters' profits. There has therefore been increasingly an effort to evaluate the costs that would have been incurred had the original investments been "prudent." This involves making a deduction from recorded original cost, and is frequently called "prudent investment cost" to distinguish it from unadjusted original cost. If rates were set that just covered operating expenses, a competitive rate of interest on investment valued at original cost, and per annum replacement cost, there would be neither profits nor losses, provided profits were not hidden in the padded values of the original cost figures. This padding would be eliminated by reductions to arrive at "prudent investment cost" estimates.

If the "fair value of investment" were estimated on the basis of reproduction costs, the investors in the utilities would be in positions more nearly comparable with those in unregulated purely competitive enterprises. The major virtue of the reproduction cost or "present value" method is that in determining value it takes into account changing price levels. When prices go up in boom periods, the reproduction cost method leads to a higher value because of rising reproduction costs; whereas in the original cost method no such change in value would be made. Again in falling price periods, the reproduction cost method makes an adjustment for price changes downward, whereas original cost (or prudent investment cost) does not. This means that investors in utilities get incomes which fluctuate more nearly with prices under the reproduction cost than under the original cost method.

But the ascertaining of reproduction costs involves serious practical difficulties. Rate hearings based on this principle are unbelievably expensive and time consuming, often running for as long as two or more years. Experts are hired by the commission and by the utility, and the technical complexity of the points covered is amazing. Out of the *mélange* of technical-politico-legal arguments comes the "informed judgment" of the commission, by that time usually out of date in view of the time consumed in the hearings. Moreover, frequently part of the property of the utility is more or less obsolete. What is its reproduction cost? The cost of reproducing the obsolete equipment, or the cost of producing improved equipment that is now available to do the work? Only the second would bring results analogous to adjustments in purely competitive industry. But if the intent is to estimate costs of reproducing on the most efficient basis, evaluation is exceedingly difficult and inaccu-

rate since there is no competitive market in which the utility buys these productive agents and there is no actual exchange on which to base estimates.

The conclusion must be that neither valuation method is at all satisfactory. Typically commissions have used some vague combination of these methods in what has been sarcastically termed by some economists the "trance" method, to indicate the absence of any recognized principles of judgment and the almost mystical manner in which commissions often seem to reach their conclusions. Outside the courts, sentiment appears to be growing in favor of the "prudent investment cost" basis, though in the courts any discernible trend seems to be toward greater emphasis on reproduction costs. Utility owners of course tend to press for original costs when these are higher, and vice versa when reproduction costs are higher.

Estimating operating expenses

Obviously, before any return can be earned on a "fair value of investment" operating expenses must be covered. These are what we have usually termed "variable costs." In general the attitude of courts and commissions has been to take operating costs as given by the utility, make allowance to cover them out of income, and proceed to the problem of return on investment. Especially has this been true of wage rates. Only in rare cases have commissions or courts enforced reductions in operating expenses where they appeared inefficiently high. The courts have been very critical of commission "management" in cases where commissions have attempted to enforce such reductions in operating costs, on the ground that it is not the function of commissions to "manage" private enterprise, but only "to see that the public interest is protected."

The importance of demand elasticity in the setting of rates

The rate set on public utility services will largely determine the quantity consumed and hence the level of output at which the utility will operate. Costs depend in part on that output, and even fixed costs per unit of service depend on the quantity of services provided. In order effectively to set rates that would cover costs, however these costs may be estimated, it would therefore be necessary to take demand elasticity into account. Commissions and utility owners have probably underestimated the degree of responsiveness of consumers to rate changes. They have proceeded as if a certain quantity of the service would be sold at any rate within a wide range; that is, they have tacitly assumed perfectly inelastic demand through these price ranges. The result has been a tendency to set rates higher than would otherwise be the case, sometimes even bringing lower profits for investors in the utilities than they could obtain if rates were reduced. Recently both utility owners and commissions have become more aware of the importance of taking demand conditions into account.

Joint products and differential rates

As if the difficulties that we have discussed were not enough, commissions are faced by yet another problem when, as is usually the case, different services are jointly produced by the same company. The electric light and power industry illustrates this situation. The electric utility may provide current for both domestic and commercial lighting and power, for urban and rural users, in mid-day and at night. Each of these groups of consumers may be distinguished from others and the electricity it consumes may be regarded as a distinct commodity. Moreover, rates to the same group of consumers may vary according to the amounts of the service used. Left to itself the utility would try to differentiate rates to each group for each use in such a way as to maximize profits. Opportunities for monopoly exploitation in the setting of these different rates would be very great.

The regulating commission is faced with the problem of arriving at a pattern of rates consonant with the public welfare. In its vague adherence to the goal of covering costs it is still faced with the difficult decision as to what proportion of the overhead costs that cannot be clearly separated should be covered from each of the services. In searching for a solution there has been some groping toward the consumers' "ideal" of relating the marginal costs of the various services to their various rates;² but since a consistent following of this criterion would rarely result in a total income just covering costs it does not solve even the theoretical problem for the commission. More typically commissions have simply followed the lead of the pattern of rates suggested by the utility owners, studying general level of rates perhaps, but giving slight attention to the meanings of differentials. The result is a rate structure based on "what the traffic will bear," which is quite different from what would be in the public interest.

Taking into account the many theoretical and practical complications of rate making, economists have recently shown increasing support for rate making based roughly on a twofold cost basis. One part of the rate would be based on the special costs of serving the consumer group in question. This charge is often called an "energy charge," as it is closely connected with the delivery of the actual energy. The second part of the rate would be an "initial" charge covering all other costs, allocated among various consumers in the most equitable fashion possible. It is pointed out that this approach to rate setting would have several important advantages. It is consistent with the general goal of finding a rate structure that will bring in revenues just sufficient to cover total costs. Since the energy charge could be based fundamentally on marginal cost, this plan would also make possible reduction of rates to any consumer group where demand was sufficiently elastic to merit such reduction.

Since it is frequently very difficult to distinguish in practice between

² See Chapters 11 and 13 and the appendix to this chapter.

rate differentials consonant with the public welfare and those that are not, it is hardly surprising to find that misuse of the theoretical analysis on the basis of which differential rates are allegedly justified is common. Intricate rate-cost analysis has been employed to rationalize and lend apparent support to differential rates in fact determined largely on the basis of "what the traffic will bear" for different classes of users. Large buyers, as the big industrial users of electricity, have the alternative of setting up their own power plants and generating their own electricity. They can therefore exert a pressure on the utility to lower their rates, while the thousands of individual consumers are relatively helpless. Even when such misuse of theoretical analysis is not clear, the elaborate discussions of joint costs in relation to rate making have created an illusion of a far more careful analytical foundation for rate setting than has ever existed in actual practice.

Evaluation of public utility rate regulation

The theoretical criteria on which public utility commissions base their procedures in rate setting add up approximately to an effort to make available to utility investors returns equivalent to what they would get on investments in competitive industries, making some adjustments in rates on different products and services according to consumer demands. Accepting either prudent investment or reproduction costs as satisfactory bases of cost estimation, what may we now say concerning the practice of utility commissions?

1. The "fair rate of return" which the commissions permit has been consistently higher than the returns generally available in other investments of comparative risk. If it just equaled these other returns it would be essentially the "implicit costs" of interest on owner-investment that we included in our theoretical analysis of costs. Since it is higher it includes an element of profits.

2. The complexity of the problem of cost estimating has given utility owners opportunity to exert pressures in the courts that have frequently led commissions to estimates higher than were justified by actual conditions whatever the theoretical basis of their procedures.

3. There has been general failure to recognize the importance of demand elasticity and the opportunities to increase sales by lowering rates.

4. The criterion of covering costs, however vaguely formulated in the minds of commission members and court judges, would be somewhat more successful if it were not for the fact that utilities typically sell several different or at least distinguishable services that have many costs in common, but at rates that may be different. Which rates should be higher, which lower, and how should costs be allocated as between these different services? In attempting to answer this question the commissions have, again vaguely, felt toward another criterion. Efforts of utilities to get rates according to "what the traffic will bear" have been rationalized into criteria of allocating resources of the utility into different

services according to marginal costs and consumer preferences. The result has been somewhat haphazard, but in recent years thinking on these matters has been leading to increased understanding.

5. A partially separate factor that many economists have emphasized is the need to take into account in rate making the effect of such rates on business cycle movements. Public utility rates have varied little through booms and depressions, though other prices and incomes have risen and fallen violently. Since utility rates are important charges their high levels during depressions are widely considered a serious impediment to recovery and a burden on private consumers; in boom periods the relatively lower utility rates are sometimes, though less often, cited as a condition conducive to a still more unhealthy boom expansion. Commissions have given virtually no attention at all to such considerations.

Regulation of Utility Holding Companies

The public utilities have long made extensive use of the holding-company device. Hundreds of operating companies have been drawn together in great holding-company systems, with control pyramided in purely financial companies. In a few cases, such as the one pictured by Figure 26—2, these intercorporate relations have reached almost unbelievable complexity, the top holding companies having no conceivable justification by the test of facilitating operating efficiency. Although public control of utility security issues and financial practices is of long legal standing, until very recently such control has commonly meant nothing more than unenforced general regulations of accounting methods and partially effective control of corporate capitalizations and mergers. Even had there been a strong desire to do so, it would have been impossible for the various states to regulate intercorporate relations of utilities effectively once the holding company became widespread, reaching across state boundaries.

In keeping with other similar "New Deal" legislation, the so-called Public Utility Holding Company Act was passed by Congress in 1935, giving the Securities and Exchange Commission wide powers over public utility corporate structures and intercorporate relations and instructing the Commission to require all utility holding companies to "integrate" their structures. Under this act the SEC has carried on detailed investigations of various major utility holding companies with a view to seeing whether holding company relations were justified or whether breakup of the holding company structure into separate companies would be in the public interest. Thus far the Commission has been able to deal thoroughly with only a few of the largest holding company structures, but has required drastic "integration" in most cases investigated, finding that existing financial holding company structures could not be justified on any grounds of operating efficiency. The effect of such regulation is on investors, directly through the effect on corporate financial structures, and on consumers, through decreased returns to security-holders and through increased efficiency of operation. Lastly, the Commission

appears to have made substantial strides in regularizing accounting and internal financial practices of utilities in certain cases.

Government Regulation and Government Ownership

The history of public utility regulation in the United States over the last half century has certainly been far from impressive; yet it has not been too disheartening. The preoccupation of regulation with legalistic considerations to the great exclusion of economic analysis, its excessive costs and delays, its rigidities, and its stumbling lack of apparent purpose beyond the settlement of individual disputes stand out with uncomfortable prominence. Yet the period has seen an enormous growth in the output of public utility services, an improvement in their qualities, some decreases in their costs, and the development of a complex regulatory machinery over a fairly short period of time.

Out of the history of this period come certain fairly obvious conclusions, but unfortunately the majority of important policy questions remain to be answered. That the trend toward closer regulation must continue seems certain; it is the exact direction and nature of this regulation that is a matter of doubt. One alternative is complete government ownership and operation, a course already explored with mixed results; another is renovated regulation. Much greater emphasis on maximization of consumption through experimental rate reductions and a more "economic" approach to the problem of differential rates would be a part of either program. Renovated regulation would involve continuation and improvement of the present program of the Security and Exchange Commission in its control over utility securities.

The merits of public versus private ownership of utilities have been debated emotionally, endlessly, and largely fruitlessly for many years. On the side of private ownership and operation are cited the long-standing American belief in private enterprise, its alleged efficiency and aliveness as compared with government operation, its record of rate reductions and greatly extended use over past years, the advantages of far-flung private utility systems over small local publicly owned units, the political dangers of increasingly centralized government control over economic life. On the side of public ownership are listed the arguments that regulation of privately owned utilities has been costly, cumbersome, and ineffective, that publicly owned plants would be free from legal restraints against experimental rate and service policies intended to maximize consumption and public welfare, that costs of public operation would usually be lower than for private operation since interest rates on government bonds are lower than utility "fair returns" and since fewer and lower salaried employees would often prove sufficient, and in general that public operation is clearly the logical conclusion of the admitted inability of the utility industries to function competitively so as to more or less regulate themselves.

There is no clear-cut answer to this problem, especially since even quantitative data are insufficient to give an answer aside from broader

political and philosophical aspects of the dilemma. The great advantage of public operation would be the ability to cut away from the present mass of legal technicalities and court control that apparently precludes an economic approach to rate making, and to undertake vigorous experimental policies aimed at tapping the apparently highly elastic demand for some types of utility services and in general at maximization of consumption rather than "fair return" to a small body of investors. In view of the "essential" nature of utility services, this shift in emphasis seems vitally important. On the other hand, municipal profits on utility services may be regarded as a convenient and desirable source of funds for local governments. How much weight should be attached to the alleged inefficiency of government operation of any "business" enterprise is a matter of personal opinion.

Government ownership and operation of utilities has varied from utility to utility and from one period to another. Water works are commonly publicly operated; gas, electricity, and local transit are less often under government ownership; the other public utilities and "quasi-utilities" are almost completely under private ownership. In the fields of gas and electricity, government ownership has been confined predominantly to smaller municipalities, although in the last decades some larger cities have been added to the list. Federal grants under PWA and other federal depression assistance have clearly played an important role in the recent expansion of municipal relative to private ownership, especially in the electric field where public ownership has shown rapid strides in recent years. However, privately owned plants still greatly overshadow public ones in both electricity and gas. For example, municipally owned electric generating plants by 1940 still showed less than 3,000,000 kilowatt hours rated capacity as compared with over 33,000,000 for privately owned plants. Virtually all the big plants are private.

With the plans for the unprecedented Tennessee Valley Authority, the "New Deal" federal government in 1933 stepped into the electric power field. With its gigantic generators and transmission facilities, TVA has found it possible to supply the needs of the territory for many miles in all directions at exceptionally low costs. Set up to undertake a program of flood control, the improvement of navigation and the production of nitrates, as well as the production and sale of electricity, TVA has found it impossible precisely to allocate its costs of generating and selling current. Many critics argue that in fact it operates at a loss although its official reports indicate sizeable profits even after setting aside funds to redeem its bonds in due course. In any case, its low rate policy has produced amazing evidence of the elasticity of consumer demand for electricity when drastic rate reductions are undertaken. Since this policy has been coupled with educational propaganda and the low cost sale of many electric appliances in the territory served, it is not entirely clear whether the increased sales have resulted from the elasticity of demand under lower prices or from increased demand, although there is reason to believe the former is of great importance. In spite of the bitter internal

and external conflicts over its policies, it has become evident that on the whole TVA has thus far proved a major success, both economically in providing a greatly increased volume of service at reduced rates and socially in the carefully planned program of land use and rural and urban rehabilitation of the surrounding area. Perhaps equally important has been the educational effect of the TVA experience with drastic rate reductions on rates and rate attitudes throughout the country. Whether the other great public power projects undertaken on a similar basis will prove equally successful remains to be seen. The elimination of political and personal frictions evident in TVA operation should permit more successful operation than has thus far been possible.

Other evidence on government operation of utilities in this country is inconclusive. Most municipally owned utilities have been moderately successful, though seriously handicapped by their small size and inability to attain real operating efficiency. In some cases efficiency has been encouraged by inter-community competition to attract population by offering better and cheaper utility services; in other cases graft has been shocking and rates high. The experience of the nation's railroads under federal governmental operation during the First World War was hardly a happy one, although it is not certain that the situation would have been a great deal better under private operation, in view of the tremendous stress placed on the transportation system by the war effort. In any case, such an experience, undertaken without preparation and running over an unprecedentedly strained period, can hardly be taken as conclusive evidence of the efficiency of government operation of the rails.

Whatever the course chosen, be it more effective regulation or government operation, the crucial consideration is the same—the need to break through established rigidities and traditions and push strongly toward lower costs, lower rates, and better service. In the end, aside from questions of broader social and philosophical nature, the question of which course of action is preferable resolves primarily into the question of which can best attain this general aim.

The Public Utility Approach and Antitrust Policy

The essence of the public utility approach to the control of business activity is in the enforcing of obligations on a business to act in a manner consistent with the public interest. This involves the obligation to serve whoever may seek and be willing to pay for its services, without discrimination between customers; it also involves the obligation to charge prices that will bring only a fair return and not exorbitant profits. The enforcement of these obligations would be unnecessary in an industry characterized by active competition; competition itself sets a framework of economic action in which the general public is protected. Where competition is absent, and where it is not possible to create a situation in which competition will become active, there is no such guardian of the public welfare unless the State intervenes directly. This the State does and has long done when firms are granted the exclu-

sive right to operate under a government franchise. There are many marginal cases in which the State has not granted such a right but in which government has seen fit to intervene to protect the public.

The public utility approach to the control of business activity is usually regarded as a policy to be employed in cases where there is no possibility of competition and an antitrust program would be inapplicable; but it is also a possible supplement and support to antitrust action. Here we shall consider two such potentialities: (1) the public utility approach in dealing with bottlenecks in marketing, and (2) the public utility approach in policy relating to patents. This is just a beginning. The whole range of problems involved in the coordinating and dovetailing of public utility and antitrust policy is only now commanding the careful attention of qualified students of economic structure and processes.

The regulation of marketing bottlenecks

Many of the activities carried on or depended upon by the various firms in an industry are more efficiently provided on a large scale for all the firms together. To a considerable extent this is true of technological research, and of the providing of market information. Trade associations have become agencies for the collection of considerable market information, and the federal government has also contributed here. Specialized transportation and terminal storage facilities, and facilities in big central markets are also frequently provided for more efficiently on a large scale. Under these circumstances there arise "natural" bottlenecks in marketing, which sometimes become the basis of centralization of control in a few big business enterprises at other stages of production as well.

In studying "foundations of power" in Chapter 26 we described briefly the way in which oil pipe lines have been used as a basis of monopolistic control in oil refining, the way in which shipping and storage facilities gave two large companies control over the production and sale of bananas, the way in which controls over central markets have been used as a basis of oligopoly power in meat packing. All of these provided a relatively secure basis for the building up of exclusive positions by small numbers of large firms. To a considerable extent the advantages of these large firms rested not on advantages of bigness in processing, but rather in marketing. The importance of these bottlenecks in the marketing of goods has received increasing attention in recent years; for example, the Interstate Commerce Commission is now playing a more active role in the regulation of the oil pipe lines. Insofar as such marketing services are more efficiently performed on a very large scale, our analysis seems to indicate that here is a sector of economic life in which careful government regulation may be especially needed. Two basic principles have been suggested by a number of economists. One of these is the complete separation of the ownership and control of the marketing function, such as the pipe lines or the livestock exchanges, from ownership and control of the processing businesses. The second is the treatment

of these marketing bottlenecks along the lines developed in dealing with public utilities. Such a policy could make these facilities available to all on equal terms, it could protect the public against unreasonable charges, and it could prevent the use of these bottlenecks as means to the establishment of monopolistic power in other stages of production. The difficulties of carrying out such a program are the difficulties already encountered in the regulation of public utilities, and in addition the problem of deciding in which cases this type of governmental control is desirable and in how much detail government should exert its power.

The public utility approach to patents

One of the most important bases of monopolistic power is the manipulation of patents. The issues involved in the formulation of a satisfactory patent law are many and complex. Here we shall consider only one suggestion which has been made recently and which offers interesting possibilities.³ Patents may be regarded as a kind of franchise; in fact they have been so regarded in a number of cases in the Supreme Court.⁴ They involve the granting of an exclusive right to use a certain process or to produce a particular good. In granting such an exclusive privilege in public utility franchises, government also imposes certain obligations, and undertakes to supervise and regulate the activities of the concerns involved in order to protect the public interest. In granting the exclusive rights associated with patents it has failed to apply the associated controls.

Whereas the manner of exercise of all other species of property affected with a public interest has long been recognized not only to be properly a matter of public concern but to require, on account of a peculiar susceptibility to abuse, special regulation, latterly reinforced by administrative supervision, the manner of exercise of patent franchises has continued to be regarded as nobody's business—except that of the patentee, and this policy has persisted notwithstanding the fact that the susceptibility to abuse arises in the one case as in the other from the same root cause, the monopolistic privileges conferred by the franchise.⁵

Recent studies of the structure and practices of American industry, made by the Temporary National Economic Committee, reveal in startling clarity the extent to which a patent system designed to encourage technological innovation has been perverted to the interests of particular groups who have thus built up monopolistic empires and even blocked much technological progress. The patent system has grown up as a major obstacle to the effective operation of an antitrust program, in fact of any program designed to further abundance and progress.

Confusion, cross-purposes, and mutual frustration between antitrust and patent law are shown to have suffered the erection of utterly needless obstacles

³ See the article "Present Position and Prospects of Antitrust Policy" by Myron W. Watkins in the June 1942 Supplement to the *American Economic Review*.

⁴ *Bloomer v. McQuewan*, 14 Howard 539 (1852); *Chaffee v. Boston Belting Co.*, 22 Howard 223 (1859); *Seymour v. Osborne*, 11 Wall 516 (1870); *Mitchell v. Hawley*, 10 Wall 548 (1872).

⁵ Watkins, *op. cit.*, p. 100.

to free enterprise and effective competition which even a better implemented and more zealously prosecuted antitrust policy could scarcely have circumvented. Yet those two elements of public economic policy which have actually been permitted to develop antithetic tendencies clearly might have been so shaped that they would supplement and reinforce one another.⁶

The application of the public utility approach to the treatment of patents would require detailed study of many aspects of the problem. Unless we wanted to regulate most of American industry in detail, the commission delegated with protecting the public interest in the exploitation of patents would have to operate in a manner quite different from the present utility commissions regulating the rates charged for services. What this might mean in practical detail is beyond the scope of the present discussion. But application of the principle would require that the patent holder should make the services associated with the patent available to all without discrimination and on terms that brought him only "reasonable" returns. It would eliminate the use of patents as means for obtaining over long periods huge monopolistic profits. It would bolster to a highly significant degree the implementing of antitrust policy.

Contrast between the "public utility approach" and the NIRA approach to the control of industry

There is considerable confusion in contemporary thought between the "public utility approach" and what may be termed the "NIRA approach" to the regulation of business practices. They are different in two important respects: in basic purpose, and in method of operation. The NIRA approach, carried out more recently in the coal, milk, and other agricultural industries, has been undertaken primarily to "help" those in the particular industries involved. It has been associated with the "income problem" discussed in the last chapter, and with efforts to stabilize conditions in an industry. The "public utility" approach has been geared, with whatever imperfections of execution, to the protection of the general public against the exploitation of monopolistic power. The method of the NIRA is government participation in fostering and supporting collaborative action among the members of an industry. The method of treatment of public utilities has been regulation by a commission that consults with the members of an industry but is legally a body to control and regulate the activities of private businessmen. Both are subject to abuses. There may be sectors of the American economy in which the method of NIRA can and should be used; but in advocating such a policy it is important (1) to recognize that it may easily be utilized to bolster monopolistic practices and cut the ground from under antitrust policy, and (2) to re-evaluate criteria upon which the program is based, giving full consideration to the protection of the public interest in promoting full use of resources, efficient allocation of resources among different channels of production, and progress toward greater abundance.

⁶ *Ibid.*, p. 102.

Appendix to Chapter 29

In those cases in which prices charged by private monopolists are publicly regulated (in decreasing cost industries), and in instances of

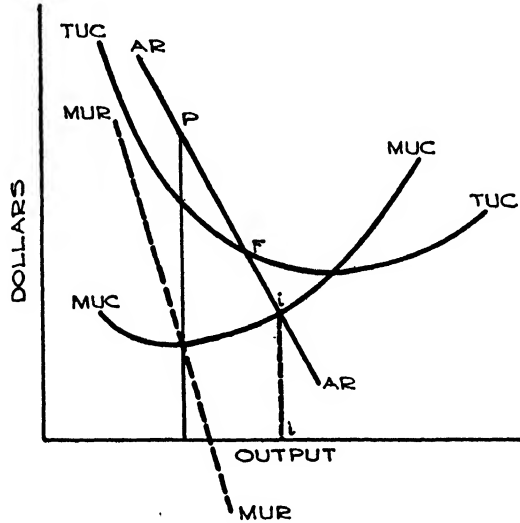


Fig. 29—A. Regulation at "ideal" rate would involve loss.

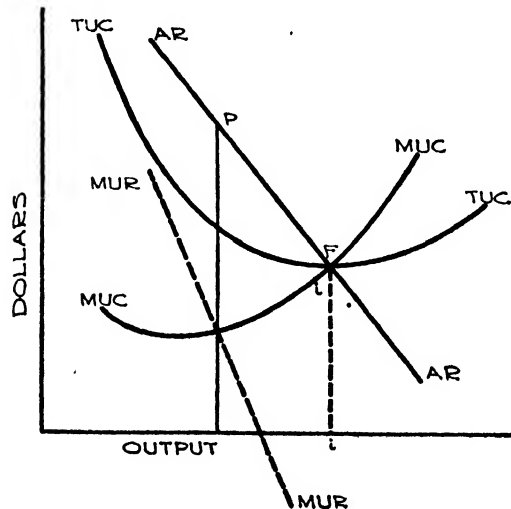


Fig. 29—B. Regulation at "ideal" rate would just cover costs.

direct provision of goods and services by the government, a real theoretical problem arises as to what rate or price would be most consistent with the public welfare. Here we shall undertake a more precise theoretical analysis than that in the body of the chapter, though many of the difficult issues will be omitted. That the analysis presented here

underlies much of the immediately preceding discussion should be evident.

The realization of the "consumers' ideal" that output (and price) should be regulated at the level at which marginal unit cost equals selling price (see Chapters 11, 12, 13) would have various results with regard to the short-run profits and the losses of a simple monopolist. One of two situations will usually arise; if the point at which marginal unit cost equals selling price is at a price which is less than the associated total unit cost, the concern will be losing money; if it is at a price greater than the associated total unit cost the concern will be making a profit.

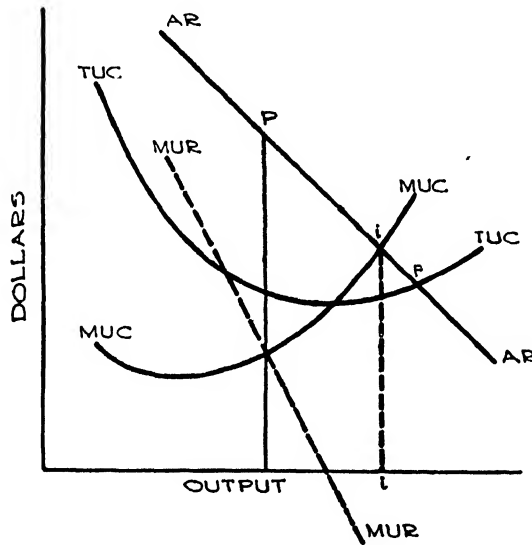


Fig. 29—C. Regulation at "ideal" rate would involve profit.

Only in the exceptional case, in which average revenue intersects total unit cost at the lowest point, will total costs and total revenue be equal, profits zero. These three situations are illustrated graphically in Figures 29—A, B, and C, where the "consumers' ideal" adjustment is indicated by the dotted line *ii*. In Figure 29—A, the "consumers' ideal" price would result in a loss to the concern, in Figure 29—B, costs would just be covered, and in Figure 29—C profits would be obtained.

The appearance of persisting profits or losses in the "consumers' ideal" adjustment poses a problem to the policy maker attempting to set up a working basis for adjusting rates. If government regulated price at this "ideal" in an enterprise in which the "ideal" adjustment would involve operating indefinitely at a loss, it would have to subsidize the firm in some way; otherwise the latter would eventually cease to function. On the other hand, the owners of the enterprise making a profit at the "ideal" adjustment are getting excess incomes that can be taxed away. It is exactly this type of adjustment through subsidies and taxes that is recom-

mended by some of the welfare economists, though its practical application would be exceedingly difficult. There is the alternative of compromising with consumer interests by using some other criterion for regulating the price the monopoly is permitted to charge.

In government efforts to regulate prices charged by public utilities for their services, the emphasis has generally been on establishing a rate that would just cover costs, no more and no less, rather than on the "consumers' ideal" rate we have just analyzed.⁷ In most of the literature of the subject this "fair" rate is usually described as a rate that will cover "costs" (based on a "fair value of investment") plus a "fair return" on owner investment. "Costs," as we have defined them in our theoretical analysis, however, have included a competitive return to owner as well as lender investors. A "fair" rate (price) would then be a rate that would just cover "costs" as we have defined them. It would be at the level at which average revenue (selling price) equals total unit cost. There would be no need either to subsidize or to tax the utility if the regulated rate were the "fair" in contrast to the "consumers' ideal" rate.

The effect on resource allocation of setting the "fair" in contrast to the "consumers' ideal" rate will depend on the particular relationship existing between costs and average revenue. The comparison is presented graphically in Figures 29—A, B, and C. In the first case, demand (average revenue) is not great enough so that a price equal to or greater than cost could be obtained for that output at which total unit cost is lowest. The "fair" rate, f (where average revenue equals total unit cost), is at a smaller output and higher price than would be "ideal" from the consumer point of view. Additional resources which consumers would prefer to see employed here are not so employed. In the last case, on the other hand, demand (average revenue) is greater. The "fair" rate, f , is, as always, where average revenue (selling price) equals total unit cost; but in this case the rate set is lower, the output greater than that which would be "ideal" from the consumer point of view. Resources that consumers would prefer to see employed elsewhere are brought into use here instead. As may be readily observed, both "fair" rate and "consumers' ideal" rate are in each case lower than the price that would be charged by the utility if it were left free to maximize profits by producing the output at which marginal unit cost and marginal revenue are equal.

⁷ The problem of rate setting on joint products has, however, introduced problems of adjusting cost estimates between the different products. In attempting to find a solution commissions have groped fumblingly toward an adjustment related to different consumer demands, and have therefore become involved in rough comparisons of marginal costs and marginal revenues of the different products and services. The result is a kind of criterion that considers overall costs and a "fair structure" of all rates to cover all costs, while the different rates are crudely related to the different marginal costs, and hence to criteria roughly approximating a "consumers' ideal" level of the different rates. The development of the distinction between the "initial" and the "energy" charge in electric rates is an expression of this compromise.

Part VI

LABOR ORGANIZATION AND LABOR MARKETS

PREAMBLE

THE "man on the street"—if it is a city street—is a wage earner. A hundred years ago he would have been running his own store or factory, or working for hire in a small shop with a few other men. Today he is more commonly one of many workers in a large establishment. A hundred years ago he and his employer would have known each other well. Today they rarely meet. A hundred years ago he would have approached his employer individually in seeking work, talking about wages, even discussing production problems. Today he is increasingly likely to be represented by a labor organization, called a "trade-union." He has become a part of the "American Labor Movement."

What is this thing that we so loosely term the "American Labor Movement"? Why was a labor movement so slow in developing in this country? Do we really have one today? Trade-union organization in America has been so pragmatic—so focused on the practical questions of wages, hours of work, restrictions of entry of outside workers to jobs under the jurisdiction of the union—that many people would say it was not really a "movement" at all. But whatever we call it, organized labor is playing an increasingly important role in American life. It is no longer just skilled carpenters and miners and garment workers who are organized, as was true two decades ago. It is also the semiskilled men putting in the bolts on the automobile assembly lines; it is the men firing the big furnaces in the steel mills. And organized labor is participating not only in wage making and the determination of working hours; it is participating in broader aspects of industrial management as well.

With the growth of organized labor comes a change of colors and forms in the broad picture of the American Economy. Here, as in the ranks of industry, is a drift from individualistic policy making toward group action. Here is another powerful force affecting the way incomes

are divided among the members of society, affecting the way not only labor but other resources as well are used, affecting the level of productive activity over large sectors of the economy. And here is another set of problems for policy makers.

What is public policy toward labor, and what will it become? We see many changes going on before our eyes. We see government legislating that employers shall pay "living wages," and supporting efforts to establish job bargaining by organized labor in place of the individual bargaining of earlier days. We also see efforts to check some of the monopolistic and restrictive policies of increasingly powerful organized labor groups. But policies are not clearly defined; they are in a state of flux. The structure of the American economy of the future will be profoundly affected by the unfolding of the labor movement itself and of public policy toward wage workers and their activities.

CHAPTER 30

The History and Philosophy of Trade-Unionism

WHEREVER there are common interests shared by a particular group in a society, there is an incentive to combine in group action to attain the common ends. Group organization appears at every turn in the modern American economy. Some groups are very cohesive, others are very loosely knit. We have a National Association of Manufacturers; we have trade associations of the producers in particular industries; we have a Farm Bureau Federation, a National Grange; we have the American Federation of Labor, the Congress of Industrial Organizations, the Railroad brotherhoods, and so on and on. Trade-unions are of many different kinds, but all are organizations of groups of workers who sell their labor services and who band together in the common interests that unite them in their working lives.

In 1894, Sidney and Beatrice Webb, famous historians of British unionism, defined a union as "a continuous association of wage earners for the purpose of maintaining or improving the conditions of their employment."¹ They later broadened this definition by substituting for the phrase "conditions of their employment" a new phrase "conditions of their working lives."² Under such a definition come many different kinds of trade-unionism, characterized by marked differences of philosophy and of strategy. Excluded from the definition, however, are sporadic temporary associations of workers where there is no continuity to the organization. Trade-unions have played an increasingly important role in labor markets since they first began to take hold in Europe a century and a half ago.

Trade-Union Purposes and Social Philosophies

There are two distinct developments in the philosophies of labor movements. The strongest and most successful is the pragmatic approach through which labor groups seek to take advantage of whatever opportunities may be open to them within an existing socio-economic order. Such is the "home-grown" variety of trade-union philosophy in America.

¹ Webb, Sidney and Beatrice, *History of Trade Unionism*, Longmans, Green & Co, New York, 1894, p. 1.

² *Ibid.*, 1920 edition.

Quite different is that philosophy of trade-unionism, provided for the most part by intellectuals, that sees in these organizations of labor groups an instrument of the ushering in of a new social order. The broad and sweeping social ideologies associated with such "revolutionary unionism" have played a relatively minor role in actual economic and social developments in the United States and England.

Trade-unionism and the collective management of scarce job opportunities

One of the most interesting and keen interpretations of trade-unionism is that which has been expounded by Selig Perlman in various books over the past fifteen years. According to Mr. Perlman, the behavior of an economic group can be explained on the basis of the kinds of economic opportunities open to it and on the attitude of the group toward these opportunities. Trade-unionism comes into its own when workers have a sense of definitely limited opportunity. He draws a contrast between the optimism and abundance-consciousness of the business entrepreneur and the scarcity-consciousness which he regards as typical of the manual worker. When the manual worker becomes convinced that he lives in a world of limited opportunity, this has a decided effect on his attitude toward group action. This attitude toward collective action is well summarized by Perlman in the following words: ³

If opportunity is believed to be limited, as in the experience of the manual worker, it then becomes the duty of the group to prevent the individual from appropriating more than his rightful share, while at the same time protecting him against oppressive bargains. *The group then asserts its collective ownership over the whole amount of opportunity*, and, having determined who are entitled to claim a share in that opportunity, undertakes to parcel it out fairly, directly, or indirectly, among its recognized members, permitting them to avail themselves of such opportunities, job or market, only on the basis of a "common rule" [such as standard wages and hours of work]. Free competition becomes a sin against one's fellows, anti-social, like a self-indulgent consumption of the stores of a beleaguered city, and obviously detrimental to the individual as well. A collective disposal of opportunity including the power to keep out undesirables, and a "common rule" in making bargains are as natural to the manual group as "laissez-faire" is to the businessman.

There are several interesting points suggested by this quotation. First, there seem to be some clear analogies between Perlman's analysis of the scarcity-conscious behavior of the laborer and the developments in agricultural control programs in recent years. An attitude of limited opportunity in the markets for agricultural products explains at least in part the acceptance of a program involving considerable collective control over individual action in agricultural marketing and production. Second, we may ask whether there is any spread of such an attitude of limited opportunity among businessmen, and how it affects any collective or group actions they may take. Agreements among businessmen to main-

³ Perlman, Selig, *A Theory of the Labor Movement*, The Macmillan Co., New York, 1928, p. 242.

tain prices, share markets, and so forth reflect both the pressures of a market in which opportunities appear to be scarce, and the positive business advantages obviously obtainable through oligopoly control. How far the sense of scarce-opportunity enters in is a matter for some speculation. Third, and most important for our immediate discussion, in summarizing the attitude of a scarcity-conscious group toward collective action Mr. Perlman has brought out the two chief types of trade-union behavior in the management of job opportunities. These are (1) the attempt to exclude newcomers and outsiders from the particular job area under the purview of the union, and (2) the attempt to set up standard regulations concerning such important matters as wages, hours of work, and conditions of employment, and to get universal compliance of all individuals with these collectively determined "common rules."

The "scarcity-consciousness" to which Mr. Perlman ascribes the growth of trade-unionism in labor markets is the basis of a pragmatic, not of a utopian revolutionary, unionism. He brings this out very clearly in criticizing intellectual socialists:

Perhaps in abstract reasoning, the wage earner might be expected to envisage the whole of the economic organization of society as the ultimate source of his scarce job opportunity; and therefore wish for a complete "Workers' control" of industry. Actually, however, the typical wage earner, when he can express himself in and through his trade union free of domination of intellectuals, who are never too bashful to do his thinking for him, seldom dreams of shouldering the risks of management. Ordinarily he traces the origin of his opportunity not much farther back than the point where it materialized in jobs, and will grasp and support only such union policies as will enable or force the employers to offer more jobs, equally available to all fellow craftsmen, and upon improved terms.⁴

Trade-unions viewed as instruments for the attaining of a new social order

Broad visions of social change, usually stimulated in the first place by intellectuals, have had relatively little effect on the development of trade-unionism in America. Some of these ideologies have, however, played a more significant role in the development of trade-unionism on the continent of Europe. The part played by such ideologies in American labor history will be reconsidered in the next section on the history of trade-unionism. Here we may ask what they are.⁵

It is convenient to classify the approaches to a "new social order" under three main headings: (1) the deterministic-revolutionary ideology of the Marxists, (2) the attitude that the ultimate value is in freedom of the individual to run his own affairs, be or select his own boss, emancipate himself from "wage-slavery," and (3) the attitude that planning in socio-economic adjustments is important in order to avoid waste and inequalities prevalent in the existing order. No one of these groups

⁴ *Ibid.*, pp. 246-247.

⁵ This classification is based on Perlman's discussion of "Intellectuals and Labor" in *The Theory of the Labor Movement*.

of ideologies can be adequately described here, but a brief suggestion as to the meaning of each will serve to give us a better perspective in viewing what trade-unionism is and what it is not.

1. The Marxist is essentially a believer in inevitable social and economic revolution, in which the mass of class-conscious propertyless workers (the "proletariat") will take over the control of the society. He is a "determinist" in regarding the proletarian revolution as inevitable. He also regards it as desirable. He argues that the capitalist system founded on the institution of private property contains within itself the seeds of its own destruction. Technical and economic evolution will create pressures too great for the system to bear. There will come an increasing cleavage between a working proletariat and the owners of industry, with the progressive disappearance of the more stable "middle class." This working proletariat will become increasingly class-conscious and must eventually hurl its mass in a great revolution against the society supported and defended by the propertied classes. The actions of workers may retard or hasten the occurrence of the inevitable events. According to the Marxists, trade-unions should therefore play the role of organizations heading up these revolutionary changes. Workers as members of trade-unions are expected to direct their efforts toward the accomplishment of their role in this inevitable revolution which will place the control of society ultimately completely in their hands. Compromises with the present system are regarded as weak, and as evidence of the corruption of labor's leaders. In unions under such opportunist leadership, the revolutionary will turn to "boring from within." Since Marxists regard the proletarian revolution as both the desired and the inevitable development of the future, practical considerations of the immediate interests of groups of workers are to be sacrificed to the broad ultimate goal. The reasons for the failure of this ideology to take hold thus far in American trade-unionism will be discussed later. It has similarly failed in other English-speaking countries; but the influence of Marxism on the development of trade-unionism in Germany and other parts of Europe is very important.

2. Associated with humanitarianism and the philosophy of the importance of each individual man, are various ideologies centering around a conception of a "higher" freedom. This is an intellectually conceived freedom which goes far beyond the freedom of the job sought by the practical trade-unionist. This "higher" freedom means the "disappearance of all authority from above, and an opportunity for everyone to participate in the total creative planning of industry."⁶ Channels to its attainment may be of various kinds, and both the ultimate goal sought and the means to it are usually quite nebulous in the minds of even the intellectuals who have developed this philosophy. Anything giving labor a greater measure of freedom from the "boss" would seem to these thinkers to be leading in the right direction. The ultimate attainment of the goal

⁶ *Ibid.*, p. 284.

may be via something so concrete as producers' cooperatives, or "producers' guilds," in which workers are also managers. It may be via a social revolution leading to the reorganization of the society as a whole. Trade-unions as associations of working men become to these intellectuals instruments through which workers may gain a greater and greater voice in the management of industry. Up to a point, this ideology may function effectively in partnership with the opportunism of pragmatic unionism; but the goal of the intellectual is much more far-reaching than is that of the ordinary trade-union member. The freedom for workers to take over at once the risks, the responsibilities, and the privileges of management of industry is not the kind of freedom the workers themselves seek.

3. An ideology which seems to most Englishmen and Americans as somewhat more "practical" is that which emphasizes efficiency in the organization of society to avoid waste and inequality. When the avenue to such improvements is visualized as one of extensive central planning of the many activities of the economy, it merges into another form of "socialism" as a goal. The "Fabian socialists" in England represent this group of thinkers. They enter politics and seek to further social policies leading to a gradual evolution in the direction of a more and more completely planned economy. Social legislation that will fix minimum wages, minimum conditions of employment, maximum hours, unemployment and old-age benefits, and so on, is encouraged by this group of intellectuals. Trade-unions are then viewed as organizations of workers who may act both through collective bargaining and through political methods to contribute to the attaining of a more efficient socio-economic order. Again, up to a certain point the policies of these social planners will be quite harmonious with those of pragmatic trade-unionism. But where the safeguarding and expansion of scarce job opportunities for particular groups of workers (the carpenters, the plumbers, the railroad trainmen, and so on) come into conflict with the broader goals of general social efficiency, the paths of these intellectuals and of the pragmatic unionists will part. They also part when labor as a whole comes *too markedly* in conflict with other economic groups, such as farmers, or even the general mass of consumers.

History of Trade-Unionism

The history of trade-unionism cannot be separated from the history of labor movements. Such movements have great complexity in the historical setting in which they occur. They reflect the socio-economic structure of the time, the philosophies of intellectual leaders, and the political pulls and tugs of various economic and social groups.

Trade-unionism had its first real beginnings in England, where industrialization was a generation ahead of similar developments on the Continent. Although at first the British labor movement was expressed in mass unrest and political demonstrations, there early developed a pragmatic job-conscious unionism. These early unions were mainly of

skilled artisans, and were concerned with the immediate day-to-day problems of the workers.

The German developments in labor organization, on the other hand, were of a broad political character, stemming from intellectual leaders of social revolution. They were built on a foundation philosophy of class solidarity. The success of the more pragmatic unionism prevalent in Great Britain led some German leaders to attempt to copy the British developments. The result was that pragmatic unionism and revolutionary philosophy frequently came into conflict. They also sometimes merged in part in the same political parties and labor organizations. Although prior to Hitler trade-unionism in Germany had become increasingly opportunistic, never has the German labor movement lost completely the class-conscious revolutionary spirit characteristic of the Marxian revolutionary socialists.

Unionism in Great Britain has spread, and includes today a majority of the working population. It has become associated with political parties to a much greater degree than has unionism in the United States. There is class solidarity in the British labor movement, but it has not accepted the revolutionary philosophy prevalent on the European continent. It is primarily from the British developments that American unionism traces its inheritance. And here again, the particular historical setting in the American scene has given the American labor movement some characteristics of its own. As a background to the understanding of the American developments, we shall first glance briefly at British experience.

The development of trade-unionism in Great Britain

In the latter half of the seventeenth century the growth of commerce in England led gradually to an increasing gulf between the manual workers who no longer owned their tools or the materials on which they worked, and the employers who were in a position to provide these things. The problems of the workers, who felt increasingly oppressed as their strategic position in the market worsened, led occasionally to conflict and always to discussion wherever workmen congregated. Sporadic organization developed in many trades, and continuously functioning associations that were essentially what we call trade-unions were definitely in the picture in England by the middle of the eighteenth century. The fortunes of these organizations were very erratic at first, and attempts to suppress them continued. But before the end of the nineteenth century English employers had very generally accepted trade-unionism.

During the latter half of the nineteenth century there had developed in Great Britain a conservative trade-unionism with a full accompaniment of most of modern trade-union methods. This trade-unionism sought to improve the positions of working men *in their jobs* by action that would modify conditions within the existing socio-economic structure. Emphasis was on recognition of the status of trade-unions and on their rights in bargaining with employers, and on giving working men

equal rights with employers under the laws. Only in exceptional cases, such as control of working hours in factories, did these unions favor legal control of conditions in labor contracts. Mutual self-help was the philosophy of these groups.

The practical conservatism with which this trade-unionism developed in England in the second half of the nineteenth century enabled it to make extraordinary advances, but it also had some serious weaknesses from the trade-unionist's point of view. These weaknesses were so serious that a new group of moderately socialistic union leaders displaced for a while the old group. The reasons for this vulnerability of the conservative leaders were several: (1) The emphasis on good behavior and responsibility, on tactics rather than on force or coercion, led the leaders of unionism to go so far as to give up the weapon of the strike. As a result, the development was arrested even before it was thoroughly ensconced for the small aristocracy of labor, not to mention the mass of working men. (2) Many of the biggest unions were heavily loaded with friendly benefit schemes—sickness insurance, superannuation insurance, burial insurance. These constituted financial burdens on the organizations that became very difficult to carry. (3) The conservative unionism of the 'fifties and 'sixties was a piecemeal business. The group unity involved was limited to each of the crafts or occupations unionized. The broader class solidarity of the workers as a whole was not a part of the picture. There was, therefore, no strong general "labor movement" under the wing of which the separate labor unions might be carried along. On the European scene, with its relatively fixed boundaries between social classes, its denial of the franchise to workers, and the permanent identification of the worker's outlook with that of his class, there was strong economic and psychological pressure to develop a cohesive labor group. (4) This lack of general group solidarity in the conservative British unionism of the mid-century was associated also with the attitude taken by the unionists toward labor's advancement by protective legislation. It was a part of the philosophy of self-reliance and mutual self-help that had been one of the strengths of the conservative unionism despite its drawbacks.

In the 'eighties a new group of vigorous union leaders arose. They were responsible for the famous dockers' strike of 1889, and the winners at the Trades Union Congress in Liverpool in 1890. These new leaders attacked the old conservative leadership for failing to push for labor's advantage even within the existing socio-economic framework. They were socialists in their aims, but as leaders they were followed primarily because they were aggressive young men with a program that promised practical advances for organized labor instead of the stagnation that had overtaken the labor movement. The victory of this new group gave the trade-union leaders new fields to conquer, and they started quickly on a campaign to organize the unskilled and semiskilled. They also turned increasingly to social legislation. The socialist philosophy was strengthened by the fact that most of the socialist leaders were devoutly religious, as were the

workers. They preached religion and socialism together in their oratory. As a result religious nonsocialist workers elected visionary socialist leaders as their representatives. The Trades Union Congress became for a time a channel for the expression of this religious socialist fervor.

The British trade-unionism of the present day reflects these influences; they are forces that have made it what it now is. Membership is almost universal. Action is for the most part opportunistic. The socialistic influence has continued into present times as a source of class solidarity; yet with all its appeal, socialistic class solidarity has not displaced the jealously guarded interests of separate occupational groups, and the philosophy of revolution has never really taken hold at all. The members of the different unions could be brought together in a sympathetic General Strike in 1926, and as a group they could finance the Labor Party in politics; but the strike and the politics have remained relatively respectable. Writing in 1928, Mr. Perlman remarked that "The heart of the British unionism is still in these jealously revered organizations that stand guard over the collective economic opportunity of each group—the jobs and the working conditions that go with the jobs." Despite important social and economic changes, this characterization of British unionism remains true today.

Checks to the development of trade-unionism in the United States

How does the history of the development of trade-unionism in the United States compare with that which we have just sketched in Great Britain? In many respects present conditions in the American labor movement parallel the developments of the end of the century in Great Britain. An essentially conservative and opportunistic unionism, largely of skilled workers, is being challenged by labor leaders playing an active part in politics and carrying on aggressive campaigns to organize the unskilled and semiskilled. Associated with this change in trade-union activities is increasing dramatization of a broad philosophy of social change allied in some cases with a philosophy of communism. What explains these developments, and the lag here as compared with developments in Great Britain? The first consideration in answering this question must be a comparison of the different social and economic conditions, as they were historically most relevant to the development of trade-unionism and of a labor movement. Any listing of the characteristics of the American scene is artificial, since these things are all interwoven one with another; but such a classification may nevertheless prove useful.

1. Until the end of the nineteenth century, the western frontier provided an escape for those who found themselves hemmed in as wage workers. While relatively few workmen in industrial trades actually joined the trek west, the possibility was ever present in the minds of the workers. Thus opportunity's door was open.

2. There were also opportunities to move upward in the social and

economic scale. Even though an individual was unable to do so, his children commonly could. An expanding economy, expanding markets, and the democracy of a new country created ever new opportunities at the top for those who had the intelligence and energy to take advantage of this situation. The inflow of new immigrants to take over the unskilled work created a push up from the bottom that increased this optimistic outlook. The way of improvement was to rise out of the working class as soon as possible; there seemed little reason to attempt to improve the position of that class in which one held transitory membership, but whose advances one might view with alarm later on.

3. Along with social and economic democracy, there was relatively early a real political democracy. While the universal franchise in Europe came only some time after the development of a mature labor movement, it preceded that development in the United States. Political democracy was an important factor preventing the development of a broad working-class solidarity.

4. The heterogeneity of the American population was for a long time a real hindrance to the development of the labor organizations. The welding together in one group of people of different languages, religions, cultural traditions—of people who brought with them some of the conflicts of their European backgrounds—this was a task difficult in the extreme. It is notable that the most successful and cohesive unions have frequently been those in which nationality and trade happened to coincide.

5. The strength of the American institution of private property is something unknown even in property-conscious Great Britain. This strength arises from the wide dispersion of property in many small holdings. The small farmer and the small businessman have had much in common with the wage-worker. The escape of the wage-worker into one of these middle-class positions was relatively easy. Since any effective action of labor involves the curtailing of the freedom of property-owners, it involves a challenge to private property. The challenge may be drastic or moderate, but in any case the employers fighting labor will bring this issue into the arena. In America they have had the support of a large mass of small property owners who shared their philosophy. Thus standing between the wage-worker and the employer, the small property owners have acted as a cushion preventing the development of sharp class cleavages; and by joining in the defense of property they have made labor a minority group.

6. The imperfect functioning of the American economy, and the inequalities of opportunity that have always been present despite the characteristics we have just noted, was generally attributed to "monopoly" or to "bigness." In attacking these things, workers and middle-class small entrepreneurs joined forces. Thus the early developments of a consciousness of scarcity of opportunity were channeled into a movement that joined middle class and workers and delayed the day when the laboring men would unite effectively in working groups to deal

with the problems of collective control of their job opportunities as *workers*.

History of trade-unionism in the United States

Despite these many checks to the development of labor organization in America, sporadic attempts at some form of unionism appeared even at the end of the eighteenth century. Attempts at action of many different kinds appeared as a series of separate events, growing directly out of the spontaneous action of groups facing varied economic situations, and reflecting also the scattered and diverse attitudes of immigrant labor from different parts of the continent of Europe. It is this diverse and uncoordinated nature of the early development of labor organization in America that makes its summarization peculiarly difficult.

It is usual to date the American labor movement from the organization in 1827 of the Mechanics' Union of Trade Associations in Philadelphia, a central organization coordinating the activities of a number of independent craft unions. This was followed by the organization of similar bodies in New York and New England. These unions were the expression not only of economic but also of political aspirations. Before the days of general manhood suffrage there were too few such places for expressing the rights and voicing the demands that property owners expressed normally as full voting citizens. For the most part, however, the unions were kept together by specific common aims, like wages and working conditions, or mutual benefit plans, such as sickness and burial insurance. The success of these local associations of trade-unions led naturally to an attempt at union on a national scale, and a National Trades' Union was formed in 1834. This body exercised considerable leadership in the agitation for the ten-hour day, which was quite generally adopted in the 'thirties. The panic of 1837 wiped out the National Trades' Union and many if not most of the local unions. Those unions that survived were not able to press for the same immediate practical ends as before, and tended to embrace instead far-reaching and idealistic schemes of reform, which indeed became so characteristic of the period that it has been described as the "hot air" period of American labor history.

It was 1866 before another really "national" labor movement arose, with the formation of the National Labor Union. This survived only a very few years. After it turned to political activity in 1870 most of the constituent trade-unions withdrew, and the National Labor Union became an empty shell. Reorganization was attempted in 1872, but the depression of the following year killed it altogether.

Meanwhile in 1869 the Knights of Labor had been organized. This at first was a secret society whose membership was limited only by the exclusion of bankers, lawyers, and persons in the liquor business. In general form it was an association of all kinds of laborers. In cities large enough to have several units or "assemblies" some of them could be and were organized on an occupational basis. In other cases previously

existing independent trade-unions, such as the cigarmakers, were chartered as separate assemblies. Thus while the Knights of Labor was a broad and all-inclusive organization, it carried within itself also the seeds of the craft type of organization that was to dominate the American labor scene for fifty years. The Knights hoped to unite all kinds of workers to help achieve education, organization, and producers' cooperatives. Despite this general purpose there was a great difference between the locals in their immediate programs of action. Wherever locals were composed largely of agrarian and middle-class people they were sympathetic to political action but not to direct labor action (such as strikes). On the other hand, many locals (cigarmakers and carpenters) thought that the national organization was interested in general objectives to the neglect of the welfare of the individual unions. Differences of opinion such as these were largely responsible for the eventual disintegration of the Knights of Labor. The event that marked its decline was its participation in the nation-wide strike, in 1886, for an eight-hour day. The strike not only failed, but it led to such violence on the part of a few members as to discredit the organizations participating even though they were not responsible.

In 1881, while the Knights of Labor was still powerful and growing more powerful, the "Federation of Organized Trades and Labor Unions of the United States and Canada" was organized. Many of the unions that participated were local assemblies of the Knights of Labor, although the two groups were later to become bitter rivals. This organization in 1886 changed its name to the American Federation of Labor. The ultimate victory of the federation was a victory for a set of policies carefully worked out from the experience of such craft unions as the cigarmakers, whose Samuel Gompers became the leading figure and perennial president of the new organization until 1924, when he died. The Knights of Labor survived until the 'nineties, but lost ground very rapidly.

The hard-won experience of Gompers, Strasser, and other leaders of the A. F. of L. indicated to them very clearly that to be successful in defending labor's interests they had to avoid dissipating their strength and diluting their enthusiasms in the pursuit of utopian dreams. If theirs was a calculated distrust of revolution, they knew that the people distrusted it with unreasoning passion and were inclined to suspect it in the labor movement; consequently they had to avoid not only its substance but its appearance. In the programs of their unions they were guided not by any fanciful charts of imaginary kingdoms beyond the horizon, but by the immediate and visible objects that would cause work to be less irksome and to yield the means for better living. They not only accepted capitalism but desired to function within it. Even the old dream of a worker-controlled economy through producer cooperation they had learned to distrust.

With businesslike opportunistic day-to-day programs they naturally adopted a decentralized form of organization, with the ultimate authority

in the hands of self-governing local unions in touch with the conditions on the job. The federation itself was for the purpose of expressing the common interest of all the locals and for preserving harmony among them. It was not long in learning that one of its serious problems was that of avoiding conflict resulting from overlapping of different unions in the same field. Dual unionism was a bugbear to be avoided at all costs.

On account of their minority consciousness and the feeling that legislation was a two-edged sword, the leaders of the A. F. of L. have taken fairly consistent stands with regard to their policies of immediate action. They have insisted on:

1. The avoidance of entangling alliances with nonlabor groups or idealistic movements such as farm groups, the cooperative movement, general panaceas;
2. Distrust of legislation as a means of securing gains, dependence primarily on negotiated agreements with employers;
3. Complete avoidance of any commitments to political parties; adoption of an opportunistic policy of rewarding friends and punishing enemies regardless of party lines.

While the A. F. of L. was dominant until recently, there were some important manifestations of unionism outside its ranks. The "Big Four" Railway Brotherhoods have steadfastly remained aloof. They were powerful enough in their own right so that they could afford to remain independent. Others stayed out of the federation because they did not agree with its program. An interesting example was the Western Federation of Miners during the 'nineties. Their members were rough and tough miners in the frontier districts of Colorado and Montana, and their employers were equally rough and tough. Disputes were usually gory, and the miners had no sympathy for the conservative piety of the Federation.

The Industrial Workers of the World was formed in 1905 out of some scattered remnants of the Western Federation of Miners and a few visionary intellectuals who had been thrown out of eastern unions as "theorists." By and large these latter wrote the extremely radical constitution, with its dreams of establishing a new order in which the workers would appropriate each industry. The actions of the union on the other hand were directed by the rank and file, who were interested in bread and butter. They sang "Pie in the Sky," not the "Internationale." They were, it is true, for the most part an unstable, transient, emotional group, temperamentally quick to violence. They were of a social class that received little sympathy from the people with whom they came in contact. Their story is scarcely that of a social menace, but of a human tragedy, though they were feared and hated at one time with a hysterical and bitter hatred.

The outbreak of World War I in 1914 found the labor movement in the doldrums. For ten years unions had merely been holding their own

in organization, and they were in the grip of depressions. Politically however their support of Wilson in 1912 had left them in a more favorable position, and the revelations by the Commission on Industrial Relations of the callous inhumanity of certain employers had practically for the first time won them widespread public sympathy. As the war progressed prices were rising, the cost of living was going up. Naturally labor wanted higher wages so that they could at least maintain their former plane of living. Meanwhile with immigration from war bound Europe practically cut off, it was easier to obtain wage increases. The long-dormant movement for the eight-hour day gained ground in many industries, particularly in those industries where there was a great influx of women workers. When the inability of the railroads and their employees to settle the eight-hour issue threatened to cripple vital transportation facilities Congress passed the Adamson Act, providing for the basic eight-hour day on railroads. This gave greater weight and respectability to demands for the eight-hour day in other industries.

At this point the United States itself entered the war. Orderly labor relations became even more vital to national defense. Far-sighted men like Carlton Parker not only began to present the idea that by attempting to understand labor and work with it in adjusting legitimate grievances and reasonable demands it was possible to forestall unreasonable action; such men were also conspicuously successful in mediating violent and spectacular disputes. The policy of the A. F. of L. was to be cooperative and avoid hot-headed action, meanwhile negotiating concessions in return. This policy proved very fruitful, so long as the war lasted.

With the close of the war however the truce ended. Labor felt that its long restraint should be rewarded. On the other hand many industries were not yet wholly reconciled to a permanent policy of dealing with unions. In the autocratic steel industry where workers as late as 1924 were still working a twelve-hour day seven days a week, with a twenty-four-hour day every two weeks when the men changed shifts, there was a violent strike in 1919. In spite of the report of the Inter-church World Movement on the facts of the case, public opinion, already outraged by the comparatively few but well publicized strikes of the wartime, turned against the labor movement. Employers joined in a nationwide campaign well timed to undermine the unions in their plants. Union membership fell to a fraction of its wartime peak. During the prosperous 'twenties with its dreams of permanent and ever-increasing prosperity it never recovered. Not until the bubble burst in 1929 did workers again become more interested in the pay check than in the stock market, and not for several years more were they in a position to start recapturing lost ground.

An old issue in the history of the A. F. of L. has been that between the craft and the industrial forms of organization—should workers organize by occupation regardless of industry, or by industry regardless of craft? Sam Gompers early settled that in his own mind; feeling that unity could be derived only from the common objectives and common problems of a

common craft. He had seen broader unions crumble from internal dissension, as each occupation group urged its own problems as the most important for the attention of the union. The federation did, however, have two industrial unions within its ranks already—the United Brewery Workers and the United Mine Workers. Those had been permitted to remain, since they were indeed powerful groups, but the policy was to admit no more industrial unions.

It was said that the division of labor made craft unionism the only appropriate kind. It set workers at separate tasks, gave them diverse and distinct problems and interests. For one union to promote the interests of all would be virtually impossible. It was a job for specialized groups who knew the special problems, the typical grievances of each group. Sidney Webb, the eminent English labor historian pointed this out very clearly in his book *Industrial Democracy*, first published in 1897.

Meanwhile industry marched on. The division of labor progressed. Tasks were divided so minutely on the assembly lines of some industries that the effect was to bring workers together in interests and type of problems rather than separate them. The plea for industrial unionism was revived.

In 1933 at the bottom of the depression American trade-unions were in a sorry state. In that year, however, the NIRA⁷ in Section 7(a) specifically guaranteed to workers the right to organize and bargain collectively without interference by employers. With no heavy handed antiunion campaign to fear, organization proceeded rapidly. It seemed that labor's golden age was at hand. It seemed to some union leaders that here was the opportunity to increase the strength of the labor movement, and to do something for what appeared to be some of the worst exploited groups by organizing previously unorganized workers. At the 1934 convention this feeling was widespread enough to lead to passage of a resolution calling upon the executive council to embark on an organizing campaign in the mass-production industries. The old heads of the craft unions, who dominated the executive council, were none too enthusiastic about this task. They felt that these new converts would have too many enthusiasms and too little judgment, and that this might lead to ill-advised strikes and unwarranted demands. They might discredit unionism with the public, bankrupt the treasury through unsuccessful strikes, thus jeopardizing funds accumulated for benefit payments, and vote out of existence long-established and successful policies before they learned to understand them. These were honest and genuine fears, even if not altogether warranted. Consequently, the organizing of mass-production workers proceeded slowly and unenthusiastically. By the time of the 1935 convention little had been done.

To many of the delegates this appeared like a betrayal. They

⁷ National Industrial Recovery Act.

demanding explanation and they demanded action. Failing to get it from the executive council, representatives of eight unions got together and formed a committee within the American Federation of Labor to work for the furtherance of unionism in the mass-production industries. This committee became known as the Committee for Industrial Organization. The organizing activities of the committee in such industries as steel, rubber, and automobiles were very successful and spectacular. The activities of this committee appeared as an increasing menace to the old-line craft unionists. The skilled aristocracy of labor saw itself threatened by these developments. Jealousies between the craft unionists and the C.I.O. leaders developed to a white heat. Failing to induce the C.I.O. to disband, the executive council of the A. F. of L. finally suspended from the Federation the unions involved. The C.I.O. then set up its own completely independent organization, later changing its name to the Congress of Industrial Organizations. Within a few years its numerical strength matched that of the Federation.

Trade-unionism in the current American scene

Today unionism in America has spread to include a proportion of the population heretofore untouched by any effective labor organization. For the first time it reaches down to workers of all degrees of skill, and has moved with dramatic success in the big mass-production industries. A liberal federal administration has supported labor organization and labor's right to collective bargaining. The last recalcitrant group of employers are being "brought into line," though most unwillingly. In general, trade-unionism is being accepted at long last as a permanent and extremely important part of the economic structure of the nation. But there are many weak spots in labor organization today, and there are many uncertainties ahead. What are the most significant elements in this current picture? Speculation concerning current trends is dangerous, but interesting. We have selected four points for consideration: (1) the significance of current internecine strife between different union groups, (2) the impact of concurrent war conditions and union organization campaigns on public attitudes and union stability, (3) the question as to whether or not there is a developing sense of class solidarity, and (4) social unrest and communistic activities within the labor movement. These four sets of considerations are obviously mutually inter-related.

1. Strife between unions seems to reflect a number of pressures both from within and from without the labor movement. The conflict between craft and industrial organization is an important one, and is directly related to technological evolution in industry. There seems little likelihood that either type of unionism will eliminate the other, but industrial unionism has been gaining relatively. This is to be expected with the technological breaking down of craft lines and the aggressive campaign to extend the front of labor organization. However, if they must bargain with trade-unions, it is much simpler for employers to deal

with the business agent of a single industrial union than with those of thirty or more craft unions. Reluctance to do so is usually due to distrust of the responsibility of C.I.O. leadership. But internecine strife is not due entirely to the inevitable economic conflict between these two bases of organization. It reflects two other factors as well. One of these is the rivalry between leaders of trade-unions in their search for personal power. This rivalry extends into sections of the C.I.O., and is not confined to C.I.O.—A. F. of L. conflicts. It is tied up with the third source of conflict, the fights between those who seek to establish a relatively conservative, job-conscious, pragmatic unionism on the industrial basis, and those who seek through the C.I.O. to further the cause of social revolution. Internecine warfare of whatever sort weakens labor's position. Whether this will result in a serious set-back to labor organization remains to be seen; such prophecy is beyond the competence of this analysis of the labor situation.

2. The defense program of 1940–41 gave labor opportunities to win further recognition and wage increases. The consolidation of gains of previous years was viewed as essential by hard-hitting union leaders. War conditions coming at this particular time in trade-union growth, exerted especially heavy pressure to engage in strikes for union recognition. At the same time, such strikes tended to evoke far more public hostility than would occur in a less critical period. It is significant, however, that the strike record in these years was remarkably moderate when matched against the record of World War I.

3. There has been much discussion as to whether we are developing in America at the present time a class-consciousness approaching that which has been known in Europe for many generations. There are many factors which might lead us in that direction. Most important is probably the closing in of markets and the juggling down of economic relations so that workers are becoming a more settled "class." In response to this, there may be an increasing awareness among workers of scarce economic opportunities. This reflects both the domestic depression of the 'thirties, and narrowing international trade. A counter-acting influence is the spread of educational opportunities which may once more open up to laboring people avenues leading to new economic horizons. Social security legislation is an expression of these changing circumstances and attitudes.

4. Communism and class-consciousness are not necessarily tied together. It may well be argued that there is a growing solidarity of working-class people, as has long existed in England, and yet that there is no real appearance of a revolutionary philosophy among these laborers. On the other hand, in this period of change and social unrest those who believe in revolutionary reconstruction of society find some listeners. There can be no doubt that such individuals have penetrated some of the local unions of unskilled and relatively casual laborers. This is not new; the I.W.W. appeared in an earlier period of social unrest. But such attitudes have not reached dimensions of any great significance.

A comparison of trade-unionism in England and the United States

We are now ready to make some summary comparisons between the history of trade-unionism as it developed in Great Britain and in the United States.

1. Trade-unionism got on its feet much later in America than in England. Though generally accepted by British employers forty years ago, it has just recently been coming into this position in the United States. There is still probably a majority of American laborers who are not union members, and a smaller proportion who do not come under the provisions of trade-union agreements.

2. In both countries successful trade-unionism developed on a relatively conservative and pragmatic basis. Objectives were direct and immediate. Unionism was job-conscious, and it first gained a foothold among skilled workers.

3. In England, despite the importance of the separate unions, the laboring class as a whole developed a solidarity in attitude and political action that has not appeared until recently, if at all, in the United States. The reasons for this are evident in the different social and economic conditions prevalent in the two countries.

4. Contemporary developments in economic and social conditions in America are setting the stage for the growth of a more class-conscious unionism, though the growth of the system of public education in part counteracts this settling down of a more mature economy.

5. Trade-unionism in England has become allied with a labor party. In America this has not been the case, partly because of the nature of our political system and partly because as a minority group labor could not turn so effectively to political action.

6. Despite sporadic attempts to swing labor into a revolutionary program, such ideologies have never really taken hold in either the United States or Great Britain.

CHAPTER 31

Participation of Organized Labor in the Control of Economic Activity

TRADE-UNIONS are playing an increasingly important role in the control of economic affairs. This influence is exerted primarily through the medium of agreements between organized labor groups and industrial management. Negotiations of organized workers with their employers are usually given the name of "collective bargaining," and the results of these negotiations take form in "collective agreements."

Collective agreements have become of increasing importance with the growing power and scope of trade-union organizations and the growing need, from the worker's point of view, of representation of his interests in dealings with the managers of large enterprises. They are commonly local agreements, covering all workers in all firms of a certain kind in a particular locality; they are sometimes regional, as in the bituminous coal agreements; there are a few important national agreements generally followed by the local units; increasingly important are the agreements negotiated with large corporations by industrial unions.

What are the attitudes of workers and employers toward collective bargaining? What are the contents of such agreements? What are the weapons of labor and management in the struggle for power and advantage? How do these things add up in predicting the future role of trade-unions in controlling industrial activity? These questions are all important, and they now challenge us.

Attitudes Toward Collective Bargaining

Attitudes and arguments of workers

The increasing interest of workers in collective bargaining is easily understood. Through such bargains they seek to improve their positions, including the wages received, the hours worked, the security of job tenure, the conditions surrounding their work; such agreements provide one of the possible avenues for keeping particular scarce job opportunities for a limited group of workers. Individual bargaining, familiar enough when the worker in a small business talks things over with his friend and neighbor the boss, is out of the question for the man on the

assembly line in the Ford Motor Company. Under individual bargaining the man who objects to being literally handcuffed to his machine cannot become insistent or he will soon be walking down the road talking to himself. Under collective bargaining a hired business agent who does not work for the company and has no fear for his job can take the matter up quietly with the personnel department with some hope of a reasonable adjustment. And under collective bargaining the individual workman may be protected from the fear that he will be at any moment displaced by someone offering to take the job for less, or even by men shipped in from other areas by an employer aiming to break down the existing balance in the labor market.

As arguments in support of their position, workers bring up a number of points. They appeal to the general public by citing instances of unscrupulous employers who bring in uninformed workers from other areas and flood the local labor markets in order to depress wages. They point out to the public and the employers that collective agreements will be of benefit to entrepreneurs in the industry because they will stabilize conditions in the labor markets and prevent one firm from gaining a competitive advantage over others by cutting wages. They argue that through settling disputes in this way employers will avoid the stoppages of work and the losses of efficiency that accompany situations in which workers are dissatisfied or feel injured, and that the general improvement of morale will greatly increase labor productivity. They also argue that they will act as recruiting agencies for qualified labor thus protecting the employer and relieving him of the need for incurring the costs and risks involved in such recruiting.

Attitudes and arguments of employers

Employers do not always look at it the same way. They do not want anyone to interfere with their powers to control their businesses in every aspect of policy making. They want to be in a position to readjust wages as they may see fit, to determine independently the conditions of work which they will provide, to hire and fire at their own discretion. In part these attitudes simply express the aims of the employer to maximize profits, and to retain a free hand in making decisions that will affect these profits. In part they are much more subtle, reflecting a desire for untrammelled power for its own sake. Frequently employers question the truth of the unionists' claims concerning the advantages of collective bargaining, and doubt the degree of responsibility of unionists in upholding their end of the bargain. The advantage of collective bargaining will appear most unimportant to employers in large enterprises that can dominate both labor markets and the markets in which they sell their products. (These are just the situations in which the workers most desire collective bargaining as a protection against powerful employers and groups of employers.)

The principal argument of employers is that freedom of management in making decisions concerning labor is essential if production is to be

efficient and if flexibility of adjustments is to be maintained. They also argue that usually collective agreements infringe upon democratic rights of individual workers to make their own decisions, and that they infringe on the property rights of management in the conducting of business.

In the past when management has accepted collective bargaining it has done so partly because trade-unions have become powerful enough to force acceptance, and partly because of the improved relations with labor and the greater production efficiency that might ensue. Recently government has stepped in and required that workers in interstate industry be permitted to bargain with employers through representatives of their own choosing.

The Contents of Collective Agreements

The earliest collective agreements were largely quite informal, confined to provisions concerning wages to be paid, the employment of union men, and in some cases hours of work. Gradually these agreements have come to include not only wages, hours of work, and the employment of union members, which remain the keystones of the agreements, but frequently provisions concerning conditions of work and safety devices, promotions, policies of hiring and firing, and enforcement of the agreement. In some cases there have also been provisions for health insurance and sickness leaves, for dismissal wages, for policies concerning the introduction of new processes that might displace workers, and even for union participation and cooperation in the determination of production policies and in making suggestions for improving efficiency in the operation of the shop.

No classification of the contents of collective agreements would be entirely satisfactory, but for convenience we may group the most commonly included provisions under three heads: (1) employment policies and job protection, (2) wages and hours, (3) enforcement of the agreement. The first essential of any collective agreement is of course the *recognition of the union* as the representative of all of the workers involved in the agreement; in the craft types of agreements this means recognition of the craft union as the representative of all the members of the craft employed by the firms involved in the agreement whether or not they are members of the union; in industrial unions it means recognition of the industrial union as the representative of all the workers in the firm who might be members of that union, whether or not particular individuals are in fact union members.

Provisions concerning employment and job protection

Provisions concerning employment policies and job protection are primarily a part of labor policy intended to maintain control over the disposal of scarce job opportunities. In part these provisions are designed to strengthen the union and protect it from destruction by antagonistic employers; in part they are primarily protection for individual workers against unfair treatment by foremen or others in the firm. There

are many different forms which such policies may take. The most important of those commonly appearing in collective agreements are: (1) closed or preferential shops, (2) specifications as to what constitutes sufficient cause for firing a man, (3) specifications as to how jobs shall be divided when work is especially slack, (4) provisions directly intended to prevent the use of methods of production that would diminish the amount of employment available to the workers represented by the union.

1. A *closed shop* exists when the employer undertakes to employ no one who is not a member of the union. Obviously an employer would oppose a closed shop unless he had already accepted the union and become thoroughly satisfied with the working relations that had been established over a period of years; and in such a case the union would have little reason to fight for a closed shop anyway. It is only in cases where unions are very strong that they are able to obtain a clause in the agreement specifying a closed shop. Closed shop provisions are very general in older agreements in the building and printing industries and among the garment workers. A *preferential shop* is probably more common than is a closed shop; this is a shop in which union workers will be given preference over nonunion workers if they have the necessary qualifications for a job.¹

2. In many agreements the *legitimate causes for firing* are carefully defined and limited, and methods of procedure are carefully laid down. The grievance committees on the railroads are well known, and the protection they provide to workers is very much valued among the men themselves. The hosiery workers have a provision that the union should be notified concerning unsatisfactory work and conduct on the part of any laborer; the union then takes the responsibility for checking on the individual involved and only after warnings have gone unheeded will he be fired.²

3. There are two principal approaches to the problem of dividing the work in periods of slack employment. One of these is the *seniority rule*; the other is *equal division of the work*, involving shorter hours all around. The railroad unions are famous for their seniority provisions, to the effect that the man longest on the job shall be the first to be promoted and the last to be laid off, and vice versa. These will be discussed more fully in Chapter 32. Outstanding examples of provisions for "equal

¹ A recent development is the "maintenance of membership agreement." An example is the agreement in the Kearny shipyards, with United States Steel. This provides that all those who are members of the union or who become members during the life of the agreement must be forced to remain so on pain of being fired, or of having collected from them union dues plus any fines the union may impose on them. It is a compromise between closed shop and open shop, since present employees who are not union members do not have to join, and nonunion men can be hired.

² *Structure of the American Economy*, p. 328.

division of the work" are those in the garment industries, the hosiery industry, in breweries, and in most bakery agreements.³

4. Attempts to check competition from other methods of production take many forms. These include provisions concerning the introduction of machines that displace labor, sometimes controlling the rate at which these innovations may be applied; such agreements may simply change the period of readjustment without actually blocking the change. They also include provisions in agreements that specify the number of men to be hired on a particular job even though changed methods may make some of these workers quite superfluous. The most notorious examples are in the building trades, especially in Chicago and San Francisco; they have been a major factor retarding the development of prefabricated housing. When carried to extremes these policies may in fact involve collusion between workers and employers in a joint program to support each other in monopolistic positions while milking the public. Their economic effects will be considered in the next chapter.

Wage and hour provisions

High wages and short hours are probably the primary goals of union policy. The provisions concerning employment and job protection might be viewed as methods for attaining and maintaining these wages and hours, though they are also significant in themselves as devices for providing security to workers. Wage and hour provisions are not easily separated. The setting of a basic working day with extra pay for overtime work is both a wage and an hour provision.

Where workers are paid by the piece, instead of by the day or hour or week, there are frequently very detailed specifications.⁴ Coal miners, who are paid according to the amount of coal mined, protect themselves by provisions for checkweighmen to be selected by the workers themselves. In the case of the garment workers a committee that includes union representatives, workers in the particular shop, and representatives of management is responsible for the adjustment of piece rates. In the automobile and rubber industries there are provisions for "negotiation on matters of production standards, speed, and wage rates."⁵

Basic in all wage and hour provisions is the insistence on a *standard rate*. All persons doing the same thing for the same period of time and under the same working conditions are to receive the same return. Exceptions to this are extremely rare. It is the fundamental basis of protection of the rate of earnings from the nibbling away that would be possible if employers were permitted to distinguish between different

³ *Ibid.*, p. 328.

⁴ ⁵ These detailed provisions are partly to protect workers against unfair treatment of individuals, partly against the "speed up" of work that pushes workers and is often associated with downward adjustment of rates per piece, and partly against unintentional discrepancies of rates where the adjustments are very complex.

⁵ *Structure American Economy*, p. 328.

workers. It is fundamental in union tactics designed to prevent the undercutting of wages by "scabs."

Provisions for enforcement

Provisions for enforcement of agreements are extremely important. The variations of detail are so great as to make discussion of them here quite out of the question. There are three major aspects of the enforcement problem: (1) interpretation of the provisions of the agreement, (2) questions as to whether the parties to the agreement are living up to their obligations, and (3) adjustment of issues related to the agreement but not specifically treated in any of its provisions.⁶ In some agreements there are elaborate provisions concerning the machinery for enforcement and the settlement of disputes. In others very little is said about this problem, and adjustments may then be settled (if they are settled) through informal contacts between union agents and representatives of management.

Labor Disputes

Union policies and weapons in collective bargaining

For the most part the general public is unaware of the activities of unions except when a strike is declared. A strike is a concerted cessation of work by a group of employees with the purpose of forcing the employer to attend to their grievances, and with the intention of returning to the job when grievances are redressed. Usually the objectives are concerned with wages, hours, working conditions, and recognition of the union in collective bargaining. The strike may be called either to obtain an improvement or to prevent an adverse change. It has been noticed that a strike, if not so prolonged as to be utterly demoralizing, gives the members a feeling of solidarity that bolsters the morale of the union. It is even a serviceable organizing technique, since a strike attracts new members, and is often used primarily as such, although specific demands are always asserted against the employer.

In industries in which trade-unionism has been long established peaceful collective bargaining may proceed quite smoothly year after year. More often it will meet at least occasional snags. The effectiveness with which a trade-union can bargain will depend upon the power that it can muster to enforce its demands. Whether it will have to use this power directly will depend in large part on the attitude of employers and on custom in the industry. Whether or not force will be resorted to will depend also on the temperament of trade-union leaders and their estimates as to the possible successful outcome of a forceful policy at a given time. Established unions seldom have strikes. When they do they fight for a vital objective. Moreover, they have already planned the strike, pre-

⁶ Hoxie, Robert Franklin, *Trade Unionism in the United States*, Appleton, 1922, p. 266.

pared for it by having available money in their treasury or by negotiating a loan from another union, and they go into the strike determined to win. These conditions would characterize any strike in such strong and established unions as the United Mine Workers of America, the Railway Brotherhoods, International Typographical Union, the International Ladies' Garment Workers. There are other reasons for lack of strikes by established unions. Employers of these union workers are used to negotiation and wish to arrive at a peaceful solution. Strikes are costly to both sides. Each side knows the strength of the other fairly accurately, so has little to gain by expensive strikes or shut-downs.

If a strike is to be successful the strikers must insure that they will not be replaced in the job by other workers. One device which they use to try to secure this result is "picketing." This means posting workers at the entrance to a plant to inform passers-by that a strike is in progress. Frequently, a few pickets will suffice to keep strikebreakers from entering the plant since most workers have deep ethical scruples against "scabbing." Sometimes mass picketing is resorted to with scores or hundreds of workers around the gates to discourage strikebreakers from entering. Whenever the situation becomes this desperate it is difficult to restrain violent conflict. The rank and file find it hard to refrain from laying violent hands on "scabs" going in to take their jobs. Sometimes employers welcome or even attempt to provoke violence in order to turn public opinion against the strikers and perhaps to make out a case for calling in troops. The cautious leadership and close discipline that this situation calls for is not always forthcoming.

During a strike, or after an unsuccessful strike, unions may seek to bring pressure on the employer through his pocketbook by inducing their members and sympathisers not to purchase his product. The hope is that they will have enough public support so that the employer on the "unfair" list will find it cheaper to yield than to suffer a continuance of the boycott. This is rarely a very effective weapon.⁷

All of these—the strike, picketing, boycotts—are aggressive weapons of labor warfare. Other measures that increase the powers of unions are of the kind that are incorporated in collective bargaining agreements, or in legislation. The union strengthens its position through definite agreements that recognize it as the bargaining agency for the workers and through closed and preferential shops that make its position secure. It gains through any policies that increase the attraction it will have for workers, as when it protects them against unjust discrimination in firing, or against speed-ups in piece-rate establishments.

⁷The opposite of the boycott is the union label. This label attached to commodities indicates that they were made by union labor under union conditions. The clothing workers, tobacco workers, printers, and others use the union label. The Label Trades Department of the A. F. of L. tries to stimulate consumer consciousness of the label among union members and others in order that friendly employers may profit by their cordial relations with the union.

Employer policies and weapons in collective bargaining

Employer policies toward labor organizations may be of four general types. They may (1) ignore the existence of the union; (2) actively fight it and attempt to destroy it or to eliminate its influence; (3) accept the fact of its existence but try to influence its policies and control or coerce it; (4) cooperate with it as far as reasonably possible. Although many employers have pursued very cooperative policies, we shall focus attention here on some of the techniques by which employers may fight unions.

If active resistance to unions is planned it may take various lines. The most subtle is "deflating" unions by anticipating grievances and by providing welfare services. In the period after the First World War this *welfare capitalism* blossomed out with profit-sharing plans, bonuses, thrift plans, group insurance, pensions, mutual benefit plans, and so on. Many of these were withdrawn when the firms ran into difficulties with the Great Depression. In particular the employees of many firms who had been induced to buy stock in the firm on easy payments very often felt after the crash that they had been cheated. Thus much of the goodwill obtained under these welfare plans during prosperity was purchased at the cost of resentment later on. But some of the more soundly conceived versions of the plans survived and were effective. A few outstanding firms had even more attractive formulas, such as stabilized employment with guaranteed minimum annual earnings. Of course not all businesses find their own sales steady enough to do this.

Welfare capitalism mismanaged can be anything but a source of goodwill. The excessive supervision exercised by some of the patriarchs of industry over the private lives of their employees is a case in point. It is much like forcing a child to eat spinach. It may be conceived in his interest, but it will not earn his goodwill.

More aggressive resistance to unions may take the form of promoting or encouraging the operation of a *Company Union*. This may be defined as any organization of workers as such that is intended to have no members and no affiliations outside the group employed by a particular firm. A company union is often successful in giving the workers the impression that their interests are taken care of without the need of an outside union. However, the officers of such a union are always vulnerable to the displeasure of their employers and usually ineffective in cases where there is any real clash of interest. There are a few cases where company unions have felt this so strongly that they have gone over bodily into the independent labor movement. The company unions have lost ground tremendously since the prohibition of employer financing or control of labor organizations written into the NIRA in 1933 and reaffirmed in the National Labor Relations Act in 1935.⁸ They have not

⁸ The prohibition applies of course only to industries in the federal jurisdiction, that is, to industries in interstate commerce.

been able to survive, on the whole, except in a "hothouse" atmosphere.

More usual and more popular has been a policy of maintaining an *open shop*. In practice this may mean either that the firm's policy is to permit hiring of nonunion workers, or that it hires *only* nonunion workers. If the latter aggressive policy is followed the employer has two problems: detecting and keeping out workers who are already members of the union, and preventing his present employees from joining. To the first end associated employers sometimes use the "*blacklist*" of men who are not to be hired, either in the form of a list more or less openly circulated, or in some more subtle form as inserting some secret giveaway phrase or symbol in an apparently good letter of recommendation. To the second end a device formerly widespread was the so-called "*yellow-dog contract*," wherein the worker agreed as a condition of employment not to join any union except perhaps a company union. The yellow-dog contract has now also been outlawed for interstate industries.

In any kind of antiunion activity considerable use has been made of private detectives or "labor spies" to keep the employer posted. On some occasions these men have been able to become officers of the union and report its every move from the inside. Ironically, this use of spies has not uncommonly resulted in the victimizing of the employer as well as the worker. The detective is hired to find trouble, and if he does not find it he makes it rather than report all tranquil and find himself out of a job.

There remain a few other measures that may be observed when the situation results in open conflict, such as the strike. The employer just as much as the union is interested in favorable public opinion and a friendly press, and is usually in a better position to get it. Often also by one means or another he is able to manipulate the situation so as to enlist the machinery of the law on his side. In the most unsavory cases this may be a mere matter of political connections. At the other extreme it is simply a matter of securing the equal protection of the laws. But there is a large intermediate zone where the law is either indefinite or says nothing that applies to the conflict. Under these circumstances either in good faith or as a matter of tactics employers resort to the injunction. An *injunction* is a court order issued in advance requiring a person or group to refrain from action that would permanently injure property. (Just what injures property is a ticklish question, which in the last analysis must depend on the judge's interpretation.) Violation of the injunction brings the violator in "contempt of court," which is punishable by fine or imprisonment. By means of the injunction, actions otherwise quite legal may be temporarily rendered illegal for the individual or group enjoined. Without great insight and restraint this may have utterly unfair results. An extreme case took place in connection with the Pullman strike of 1894 when the union leader Eugene Debs was enjoined from a long list of specific activities, and finally "from doing anything whatsoever." The granting of injunctions in labor cases has

recently been drastically restricted in federal courts, and in the courts of some states as well.

It is unfortunate that the occasions when labor relations get in the papers are the cases when things go wrong. There are many business firms in the United States whose personnel departments have extremely cordial relations with both organized and unorganized workers. They earn the respect of the unions and are willing to give a sympathetic hearing to even the more unreasonable aspirations of the workers. Continuous negotiation between management and labor on a basis of healthy mutual understanding must not be underestimated in either extent or importance. This kind of cooperative attitude on the part of management is likely to call forth a higher order of labor statesmanship than would be evident without it, and conversely is very often found as a response to enlightened labor statesmanship.

Looking Forward

It is now evident that the role of trade-unions in the control of economic activities may have many ramifications, and is of great importance. Some of the economic implications of the policies of trade-unions will be considered in the next chapter. Here we summarize briefly and take a look into possibilities with regard to the future role of trade-unions in the management of industry.

1. Trade-unionism has spread with great rapidity in recent years in the United States. The big mass-production industries have been effectively organized for the first time, and agreements have been negotiated with employers so adamant heretofore as Henry Ford and the magnates of United States Steel. Agreements are becoming more common and more important. More different types of workers are covered and the geographic areas over which agreements are concluded are widening. Jointly concluded agreements between groups of craft unions or craft and industrial unions on the one hand and employers on the other are beginning to appear. Agreements between large labor organizations and big corporations have recently been in the ascendance. Though we may expect that labor will experience setbacks in these respects, it seems clear that the underlying trend is toward an increased voice of labor in the management of a larger proportion of industry.

2. The contents of collective agreements are expanding. Participation in control over wage and hour policy is obviously participation in an extremely important part of the control of economic activity. In addition unions are having an increasing voice in employment policies, in decisions to hire and fire, in policies concerning the division of work in slack periods, in what innovations shall be introduced and how they shall be introduced, and in many other matters.

3. Trade-union leaders are taking an increasing interest and playing an increasing role in production planning that takes into account the

welfare of an entire industry. This is especially encouraged by the development of industrial unionism, which identifies the interests of the workers much more closely with the interests of the industry. Examples of these developments are easily cited. One of the oldest unions to take this broad point of view is the Amalgamated Clothing Workers, which has been an industrial union for many years, and has worked in close cooperation with employers in efforts to stabilize conditions in the industry. This union has even participated in plans for reorganization of establishments on the verge of bankruptcy. The International Ladies Garment Workers' Union presented a detailed plan for sales promotion that was adopted by the manufacturers in 1941. The United Hatters Cap and Millinery Workers have organized and supported campaigns to increase the demand for hats in order to support high wages and to increase employment opportunities for workers in the industry. "Hat week," initiated in Philadelphia by one of the locals, proved so successful that hat weeks have since been introduced by cooperative action of hat manufacturers and unions in other cities.⁹

A joint Mechanized Mining Commission has been set up in the bituminous coal industry, for cooperative study of the problems arising with mechanization. On some of the railroads worker participation in production policies has been developed through programs that would bring to the management suggestions of workers for improved efficiency in the shop; these have proved highly successful. Even before the Second World War the Steel Workers' Organizing Committee (one of the unions in the C.I.O.) made detailed suggestions for improvements in production efficiency, encouraging local unions to try to introduce such a program when they have become sufficiently well established; but as yet management in this industry has apparently shown little interest. Though programs of this sort can be carried out effectively only when a spirit of mutual trust and cooperation between workers and employers has developed, they may in turn help build up good relations between labor and management. They also lead to collusive action between labor and management groups "in restraint of trade." Such programs, both constructive and destructive, may well prove to be of increasing significance in the future.

⁹ Dunlop, John T., "The Determination of Wages," in the *American Economic Review*, Supplement, March 1942.

CHAPTER 32

The Economics of Trade-Unionism

THE role of job-conscious unionism in economic life is clearly important; but what are the effects of the policies pursued by these groups? We shall undertake a fairly detailed examination of two major groups of policies: (1) policies that directly exclude outside labor from participation in a particular occupational sphere; (2) wage and hour policies. After examining these two groups of activities in detail we shall then turn to a broad overall evaluation of the principal effects of trade-union activities on the allocation of resources, on freedom of opportunity in the selection of an occupation, on the distribution of incomes, on security and stability in economic life, and on the level of the national plane of living.

Economic Effects of Exclusions from "Scarce Job Opportunities"

Among the most important of the policies designed to keep exclusively for an "in-group" the scarce job opportunities available in particular labor markets are the following: (1) restrictions on entry into the union, associated with closed-shop agreements, (2) seniority rules, (3) jurisdictional disputes between rival unions, and (4) resistance to displacement of union workers by machines or by unskilled men.

Restrictions on entry into the union

Entry into a union may be restricted in a number of ways. The most common are (1) high initiation fees and (2) apprenticeship rules that limit the number of learners, exclude particular groups (especially women) and go far beyond insuring competence for the job. Such restrictions have been most highly developed in some of the skilled trades that have long been under stringent union controls. Restrictions on entry into a union are effective tools in controlling scarce job opportunities only when they are accompanied by closed-shop agreements or practices on the part of employers. Otherwise they would be broken down by the employment of workers outside the union.

Restrictions on entry into union membership, associated with the closed-shop agreements, have several important economic effects. Chief among these are: (1) By thus limiting the number of workers in an

occupation, those on the inside are able to get higher wages than would be possible if the trade were open to competition. This results in persisting inequalities in the incomes of laborers. Those excluded must frequently find employment where their wages are less than could be obtained in the unionized occupation. (2) By the same token these restrictions constitute an interference with the allocation of resources according to consumers' preferences. Too few workers are employed in the unionized occupations relative to the numbers employed elsewhere. (3) Incentives to greater efficiency on the part of the workers are much weakened since the workers do not feel the pressure of competition from those who are "on the outside." (4) Such restrictions facilitate the arrangement of collusive monopoly schemes between employers and workers.

Seniority rules

Seniority rules have been most effectively used by the Railroad Brotherhoods. Such regulations involve agreements with employers that promotions will be according to length of service, that the first men to be laid off in bad times will be those who have been the shortest time on the job, that the first to be taken on in improving times will be those longest associated with job and union. Although such agreements are not necessarily dependent on either closed or preferential union shops, they in fact usually coincide with such practices. The complete unionization of workers may appear because the workers themselves see it to their advantage to join the unions; this is very likely to be the case where the protection of union grievance committees is important. Where there are seniority rules, such protection against arbitrary firing becomes exceedingly important. Thus in the railroad brotherhoods the practice of a complete union shop is the automatic result of the attractiveness of union membership to the workers themselves.

The economic effects of seniority rules in unions where entry remains relatively free are in some ways different, in some ways similar, to the effects of direct restrictions on entry into a union: (1) Seniority rules do not limit the number of workers available, but rather the individuals who shall get any given jobs. This nevertheless provides a protection of high wages since insiders need not concern themselves about competition from new workers with junior ratings. (2) For the same reason, these regulations do not lead to restrictions of employment in the regulated occupations interfering with an allocation of resources consistent with consumers' preferences. (3) Seniority rules may affect efficiency in several ways. The relative security of the job may affect the general morale of workers in such a way as to increase their efficiency. These rules also have the effect of exaggerating the seriousness of being fired, since the loss in seniority status may be of great importance; for this reason workers will avoid with greater care any action that might give employers cause to fire them. On the other hand, seniority rules protect the established workers from the competition of new and younger

men; they thus remove one of the incentives to efficiency and at the same time they narrow the range of selection open to the employer. (4) By forcing employers to keep older men on their staffs, seniority rules extend the duration of the working lives of these men. In some circumstances this provides a corrective to the tendency to sluff off onto the society the older workers who in these occupations make some economic contribution and who might otherwise be supported from the public coffers during enforced idleness. These older workers would be less able to find alternative jobs than would the younger men who are excluded by the seniority rules. (5) For the same reason, seniority rules provide security for the "in-group."

Jurisdictional disputes between rival unions

Interunion rivalries have been one of the curses of the American labor movement, and they have grown increasingly important as technological innovations have broken down craft lines and as the "industrial" form of unionism has grown in competition with "craft" unionism. These disputes mean much more than a fight between the aristocracy of skilled labor and the mass of unskilled workers. They are fights for the life of the job, fights to retain jurisdiction over activities the character of which is changing with changing technologies. In their greatest intensity these disputes have frequently been between different craft unions within the A. F. of L. Commenting on the problem, Dr. Lois MacDonald remarks:

The attention of the public is frequently called to this problem by inter-union disputes which hold up building operations or some other job while carpenters quarrel with sheet-metal workers to decide which union's members shall hang hollow metal doors substituted for wooden doors once hung without question by carpenters.¹

Attempts have been made to adjust such disputes with the A. F. of L. by the setting up of departments to deal with them, but the powers of these departments are practically nil and they have accomplished little.

Jurisdictional disputes between unions have some very unfortunate results. First, they frequently lead to strikes in which the employer is helpless to settle the issue. Each of the rival unions thus exerts pressure on him to assign the tasks under dispute to its members. There is much time wasted, efficient organization of productive activities is hampered, and production costs are raised. The final decision is likely to depend on which of the unions is the more effective power group rather than on which would be the more economically reasonable assignment of tasks. Unions pursuing these policies are sometimes very shortsighted in fighting instead of adjusting to the stream of technological change. Nevertheless, in the immediate objective of keeping for union workers all job opportunities possible, it is a logical position to take.

¹ MacDonald, Lois, *Labor Problems and the American Scene*, Harper and Brothers, New York, 1938, p. 441.

The economic effects of jurisdictional disputes are evident: (1) They lead to waste and retard advances in raising planes of living. (2) They obviously lead to malallocation of productive resources. (3) They create rigidities in wage structures and contribute to the existing inequalities of opportunity to different workers. (4) They provide a temporary (usually illusory) security for the successful fighters, frequently at the sacrifice of ultimate greater loss. Those labor leaders who have shown the greatest vision in their labor statesmanship have ceaselessly decried these unfortunate policies and worried over the very real problems out of which they arise.

Resistance to displacement of union workers by machines or by unskilled men

Closely related to the intense rivalries that appear as the areas of union jurisdiction become blurred are the efforts of trade-unionists to retain for their members skilled jobs that are threatened by technological innovations making possible the substitution of machines and unskilled labor. There are frequently groups of skilled workers whose vested interests in their skills are washed away. The effort to hang on is as inevitable as it is sometimes desperate. Sometimes labor leaders see the drift of events, and move into the front lines to secure the first place in the new setup. Leaders of the teamsters' union in New York saw the handwriting on the wall when a generation ago the horse and wagon was replaced by the "horseless carriage," and they trained their members as truck drivers, who then stepped into the new jobs that would otherwise have displaced them completely. But labor leaders and workers are not always as farsighted as this, and frequently the readjustment is much more difficult to make. Frequently old ways of doing things have been defended to the last ditch. We have already mentioned the fact that unions are commonly able to secure agreements with employers that provide that a certain number of skilled workers will be employed on a job even though there is no longer any use for so many men. And sympathetic strikes of related labor groups may prevent completely the introduction of new labor-saving equipment where the displaced group would not otherwise have been able to defend its traditional position. Frequently, especially in the construction industry, legal standards are set up ostensibly to protect the public but in fact to prevent the introduction of new materials and methods of production that would partially displace some of those laborers (and businessmen) at present in the industry. Many of the provisions of local building codes concerning plumbing and fireproofing are of this kind, and have proved major handicaps in the development of improved techniques in housing and of better housing. In some cases these codes contain provisions that were previously advisable but that with new processes have simply become outmoded, and interested groups now prevent their revision. The effects of these resistances to change are essentially the same as the effects of jurisdictional disputes between rival unions.

Economic Effects of Wage and Hour Policies in Trade-Union Bargaining

The wage and hour policies of trade-unions are closely interrelated. The understanding of the effects of these policies requires an intensive examination of these interrelations and of the differences between the effects of trade-union policies in markets where employers buy labor services under purely competitive conditions and those in which they are oligopsonists and simple monopsonists. In pursuing this task we shall proceed under three main headings: (1) the effects of high-wage policies in markets in which employers are purely competitive buyers of labor services, (2) the effects of high-wage policies in markets in which employers are monopsonists or oligopsonists, and (3) the effects of reductions in hours. In all of these situations we shall observe trade-unions functioning as agencies for the sale of the services of the workers who come within their jurisdictions, whether these workers are members or not. Thus we are viewing a particular kind of oligopoly, a formal oligopoly controlling the conditions of sale of labor services.

Effects of union wage agreements where buyers of labor services are purely competitive

In markets in which the buyers of labor services are purely competitive, workers will be added to the staffs of the entrepreneurs involved so long as wages are less than the marginal revenue obtained by employing an additional worker. Moreover, wages will be bid up to a point at which the quantity of labor offered equals that demanded, and in this adjustment the wages of the workers will equal the marginal revenue from this quantity of labor. Can a trade-union step into such a market and raise wages above this competitive equilibrium level?

It can do so only at the cost of some unemployment of previously employed workers, or, in periods of expansion, at the cost of a smaller increase in employment than would otherwise occur. At any higher wage, marginal revenues will be less than marginal costs and it would be economical for the employer to reduce the number of workers employed. Those who remain will then be able to get the higher wage. The kind of adjustment involved is illustrated graphically in Figure 32—1. An increase of wages from the assumed equilibrium level of about \$38 per week to the new trade-union level of about \$52 per week results in this firm in a laying off of about 90 workers. The other firms involved would respond in similar fashion, with a resulting general reduction of employment of the members of this occupation.

Thus far we have considered adjustments within individual firms as based on the initial schedules of marginal revenues, assuming given quantities of other agents to be employed. This is only a first approximation. At the higher wage there is an incentive to substitute other productive agents for some of these labor services. If there are readily available close substitutes in other productive agents, the higher union

wages will lead directly to such substitutions, and the resulting unemployment of union workers will be exaggerated. Even where such substitutions are not readily available, persistent high-wage policies are likely to stimulate the development of labor-saving devices and these may in the long run lead to extensive displacements of union workers.

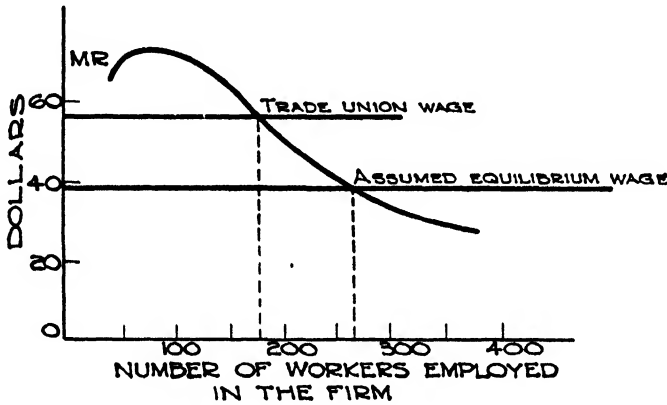


Fig. 32—1. Trade-union wage and employment: purely competitive firm.

Finally, if some of the buyers of the labor services are also purely competitive in the sale of their products, any given schedule of marginal revenues in a particular firm was based on the "going" competitive price at which these products could be sold. With the curtailment of output *in the industry as a whole*, prices will rise, and marginal revenues at each input in each firm will be larger than before. This will check somewhat the further curtailment of employment. The more inelastic the demand for the product the more quickly will the price rise, preventing further curtailment of employment. The more elastic that demand, the smaller the effect of price increases as a factor checking lay-offs. Thus when a group of laborers is hired by firms in different industries, curtailment of employment as between those industries will be very uneven. More will be laid off in industries in which demand is highly elastic, fewer in industries in which demand is inelastic.

What happens to the workers who are laid off? They may do either of two things—remain unemployed, hoping for re-employment at union wage rates, or seek work elsewhere.

If they remain idle, there is an obvious waste of resources. National real income is less than it would have been had these idle workers stayed employed. The workers may be supported by the union through unemployment benefits, but in far the largest number of cases they must either live off accumulated savings or seek public aid. One might suppose that few persons would choose to remain idle and perhaps be forced to accept public assistance in preference to seeking a job outside the union field at

lower wages, but the fact is that large numbers of union members follow exactly this course. In depression, this is explainable largely by inability to get jobs elsewhere, but a very important factor in good times and bad is the "traditional" American feeling that high prices and high wages are always better than low. Workers will remain unemployed rather than work for lower wages; businessmen will hold prices high with low sales rather than reduce prices and increase sales. Thus through the past depression, the building trades unions in most cases refused to lower union wage scales substantially from 1929 levels in spite of the fact that the average member had only about one-fourth time work and average yearly wages for union members fell to relief levels.²

Acceptance of the principle that the way to maximize volume (employment) in any particular field is to lower prices (wages) has been very slow among union leaders, and most of them have fought vigorously against any such suggestion. Although this "high wage" conviction is most strongly entrenched among union leaders, it is certainly widespread among the rank and file of union labor and is a basic explanation for the willingness of workers to remain unemployed for considerable periods rather than accept work at lower wages elsewhere.

If unemployed union members decide to seek work elsewhere, this will have the effect of lowering wages in the nonunion occupations into which they go. Suppose unemployed carpenters seek work as farm hands. This increases the supply of farm hands, forcing down the wages paid for farm labor. To the extent that the unemployed union members seek work elsewhere, they force down nonunion wages—the higher union wages for those who retain their jobs in this case come at the expense of lower wages for nonunion workers and union members working in non-union fields.

We may therefore conclude that *when a trade-union attempts to raise wages above the equilibrium level in a purely competitive labor market, it does so at the expense of some workers who lose their jobs and are forced to look for work elsewhere.* Those who remain get higher pay. Those who leave either remain idle and are supported from public funds, or they compete with other workers in other occupations, bidding down wages there; and a widening margin between the wages of the trade-union group and of workers selling their services elsewhere appears. The trade-union usually seeks to strengthen this position for the "in-group" by restrictions on entry to the trade, otherwise its position is continuously threatened by unemployed workers seeking jobs. By pushing workers out of the industry through the high-wage policy, the trade-union has of

² In the business field, the eastern railroads steadfastly refused to lower rates and fought an Interstate Commerce Commission order for reduction. Yet when the ICC finally forced them to lower passenger fares from $3\frac{1}{2}$ to $2\frac{1}{2}$ cents per mile, they found their total income far larger at the lower rates than at the old high rates. But of course there have been many cases where businesses have used a policy of low prices and large volume.

course brought about a reallocation of labor resources that is inconsistent with the realization of consumers' preferences.

This is all quite neat and unflattering to the trade-unions, but, we may ask, do workers in competitive labor markets in fact always get this equilibrium wage? In any concrete situation there may be lags which throw the adjustment temporarily out of line to either side of the equilibrium point. If a trade-union steps in at a time when wages are lagging below the competitive equilibrium, it may hasten the adjustment upward. Such effects are sometimes important, but rarely would a trade-union stop at the point at which the competitive equilibrium was reached. As soon as wages are pushed above this level they will cause a curtailment in employment. And in periods in which an industry is contracting, trade-union wage agreements are likely to increase the lag of wages above equilibrium level and therefore to hasten the laying off of workers.

Effects of union wage agreements where buyers of labor services are monopsonistic or oligopsonistic

The situation in markets in which labor services are bought by monopsonists or oligopsonists is very different from that in purely competitive buyers' markets. *Trade-unions can in this situation both increase wages and increase employment.* Moreover, their actions in this respect will result in an allocation of resources more consistent with the realization of consumers' preferences than would otherwise be the case. How do we arrive at this significant conclusion?

We may start with a hypothetical case of a trade-union dealing with a monopsonistic employer of labor. This employer is able, in the absence of the trade-union, to push down wages by restricting his inputs. He has arrived at an adjustment in which marginal cost equals marginal revenue, but this marginal cost is much greater than the wage he pays his workers. It is greater because the wage he pays depends on the number of workers he hires, and in hiring additional workers he has to pay not only the wage of the additional man but he must also raise the wages of those already employed.

Now suppose the workers in this occupation were to organize into a trade-union that bargains for them collectively. This trade-union seeks to establish through bargaining a standard wage higher than that at present paid by the entrepreneur. If it succeeds, the entrepreneur has no longer the alternatives of hiring fewer men at a lower wage or more men at a higher wage. At any wage below that set by the union he can get no workers at all. Since he cannot push wages down by restricting inputs, he will now continue to hire workers so long as the wage is less than the marginal revenue. By stepping in to bargain for the workers, the union has in effect made the wage and the marginal cost to the employer the same. Unless the wage set by the union is higher not only than the previous wage, but than the marginal revenue of the previously employed workers, the result will be an increase in employment.

This situation is illustrated graphically in Figure 32—2. The curve *EE* is the supply curve of labor services to this firm in the absence of the trade-union. Under these circumstances the entrepreneur would hire 240 workers and pay them the wage *W*. When the trade-union steps in and sets a wage much higher than this, the employer can no longer push wages down by curtailing his employment of workers. Since he must pay the union wage anyway, the cost of adding a worker is the wage; marginal cost and wage are equal. Under these circumstances he will expand employment somewhat, adding workers so long as marginal

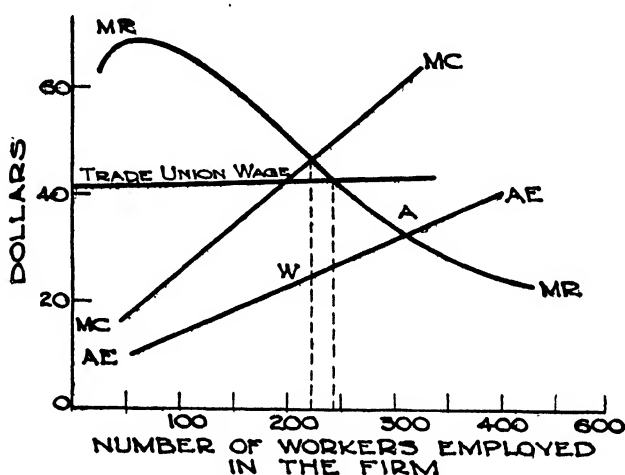


Fig. 32—2. Trade-union wage and employment: monopsonistic firm.

revenues continue to be above the wage.³ The restrictive production policy of the monopsonist will be checked by the union's setting of wage rates.

This analysis of adjustments where the employer is a simple monop-

³ How consistent is this adjustment in employment with the allocation of resources according to consumers' preferences? For the moment let us assume that the firm is selling its product under purely competitive conditions so that marginal revenues and sale value of marginal products are the same. Under these circumstances the adjustment consistent with consumers' preferences would be the one indicated by the point (A) of intersection between the average expenditure and marginal revenue curves; this is the adjustment at which all workers offering their services for a wage less than or equal to the sale value of the marginal products of those services are employed. If the trade-union were to set the wage at this level, it would result in this adjustment. Actually trade-unions usually attempt to set wages higher than this. Therefore though they may lead to an adjustment which is in the "right direction" from the consumers' point of view, it may still involve a smaller employment of labor resources here than would be consistent with consumers' preferences. Such discrepancies are of course exaggerated when employers sell their products monopolistically; but there is nothing a trade-union can do to counteract the fact that employers will adjust their inputs of labor services according to marginal revenues instead of the sale value of marginal products.

sonist is also applicable to oligopsony situations. Oligopsonist employers arrange through group action to hold wages down below the level to which they would rise were these employers to compete freely with each other. In doing this they are forced to curtail employment, since they cannot get as many workers as they would like to have at the lower wage; in fact such curtailments were the means by which they were enabled to hold the wage down. If a trade-union steps into such a situation and pursues a moderate-wage policy it may be able simultaneously to increase both wages and employment. By removing the alternative of curtailing employment and lowering wages, it leads these entrepreneurs to employ additional workers so long as their marginal revenues will be greater than the union-fixed wage.

In view of the frequency with which monopsony and oligopsony occur in labor markets, it is apparent that there is considerable scope for trade-union wage bargaining of a kind that may lead to adjustments more consistent with the interests of workers in general and of consumers as well. It is also true, however, that even in such markets a trade-union will usually set wages so high as to exclude from work some laborers capable and willing to work here for less than those actually hired. (Wages will be set above A in Figure 32—2.) From several points of view this may be regarded as a malallocation of productive resources and a maldistribution of incomes. Finally, wage increases by trade-unions selling labor services to monopsonists may sometimes be blocked by the fact that despite its monopsony position a firm is just covering costs and would be pushed out of business if the terms on which it hired labor became any less advantageous.

The effects of reducing hours worked per week on trade-union positions in wage-bargaining

Wage and hour policies are closely interrelated in their economic effects. The effects of reducing hours of work may be most clearly understood by examining first the effects on *physical* output in a hypothetical enterprise.

Suppose one hundred men to be employed, and an agreement between a union and employer reduces the hours that will be worked by each of these men from 55 to 45 per week. But the marginal product of the hundredth man may have either increased or decreased, depending on the effects of the reduction of hours on his working efficiency. If labor efficiency is greatly increased it may be that the hundredth worker providing 45 hours of work adds more to output than was previously added by the hundredth worker providing 55 hours of work. If his efficiency is not increased, the hundredth worker providing 45 hours of work will probably add less than was previously added by the hundredth worker providing 55 hours of work.

What is the relation between these physical products of the marginal laborer, and the marginal revenues? In order to simplify our analysis, we shall assume the employers to be selling their products in purely competitive markets. Under conditions of pure competition marginal revenue

and sale value of the marginal product are the same. Therefore, marginal revenue with any given number of workers in any given firm equals marginal product times selling price. Any increase of either selling price or marginal product will raise marginal revenue; any decrease in either of these things will lower marginal revenue. If selling price moves in one direction, marginal product in the other, the result will depend on which of these two movements is the stronger.

If the reduction of hours has led to no appreciable increase in the efficiency of the workers, the same number of men will produce a smaller total output than before; if this happens throughout an industry, the price of the product will then be higher than before. If the marginal product of the hundredth worker in each enterprise is assumed to be less, the marginal revenue from the hundredth worker may be either more or less. Whether it is greater or less will depend on whether or not the increased price of the product is sufficient to more than compensate for the lower marginal product. The more inelastic the demand for the product the greater the likelihood that marginal revenues may have risen. The more the marginal revenue from the hundredth laborer in each firm is raised, the stronger the position of the union in its efforts to raise wages without causing unemployment. Therefore, whether or not a policy of reducing hours will strengthen the position of the union in pressing for higher wages will depend on which of these two forces, the increase in price of the product or the decrease in marginal products, is the stronger. In either case a reduction of hours would be contrary to the interest of consumers since production here would be diminished *without* the transfer of labor resources to any alternative uses; the original output level could be maintained only by hiring additional workers, drawing them away from other occupations.

If, on the other hand, the reduction of hours results in an appreciable increase in the efficiency of the workers, there may be an increased total output of the hundred workers over what they produced when they worked longer hours. If this occurs generally throughout an industry, the price of the product will then be lower than before. Although the marginal product of the hundredth worker in each firm may have increased, each unit of the product sells at a lower price than before. The marginal revenue from the hundredth worker may be *either* greater or less than before, depending on how great the drop in the price of the product was as compared with the increase in the marginal product of the worker. Whether a policy of reducing hours will strengthen the position of the union in pressing for higher wages will depend on which of these forces is the stronger. If demand for the product was highly elastic so that the drop in price was insignificant, the workers will be in a stronger position, and the reduction in hours can be used to bolster higher wages without causing unemployment. This reduction in hours will therefore be to the interest of both consumers and laborers.

We therefore conclude: Under the assumption that employers sell their products competitively, there are four likely situations. The reduction

in hours may result in a lower total output of any given number of workers, in which case (1) if demand for the product is highly inelastic marginal revenues will rise, while (2) if demand for the product is highly elastic marginal revenues will fall. Or the reduction in hours may increase the total output of any given number of workers, in which case (3) if the demand for the product is highly inelastic marginal revenues will fall, while (4) if the demand for the product is highly elastic marginal revenues will rise. In the first and fourth of these circumstances a well-informed trade-union leader could rationally use policies of shortening hours as an adjunct of a policy to increase wages without causing unemployment of currently employed workers. In the second and third case he could not. A reduction of hours would be consistent with consumers' interests in the cases in which it leads to an increase in total output per worker, though not otherwise.

In many cases in recent years union demands for shorter hours at the same pay have really been hidden demands for higher wages.⁴ When hours of work exceed the agreed basic week, overtime is compensated at one and one-half times the regular hourly rate. A shorter basic week may not decrease actual hours of work but may instead simply increase the effective wages paid. Conversely, there are numerous instances where union members have been greatly disappointed to find that a shorter basic week actually resulted in fewer hours of work instead of the expected higher weekly pay. It is incorrect to suppose that hours of work always actually conform to the basic work week set by law or union agreement. In a large number of cases, the basic week merely determines how much of the work shall be paid at the regular rate and how much at the overtime rate.

Factors influencing trade-unions in their wage and hour policies

The decision as to what policy shall be pursued in wage and hour bargaining is a very complex one. Only to a very limited extent will the effects of these policies as we have discussed them be taken into account. Let us try to place ourselves as realistically as possible in the position of the trade-union leaders and the rank and file membership of a pragmatic job-conscious union. Immediately there are several facts that stand out:

1. All the individual members of the union would like to have higher wages for themselves. They will urge their leaders to try to get these higher wages for them.
2. Provided they can get shorter hours without a sacrifice in wages, both members and leaders will regard reductions in hours as desirable.
3. Each member of the union wants employment, and the leaders want to give their followers as great job security in this respect as possible.
4. The effects of trade-union policies on workers in other occupations

⁴ As we shall see later this has been a primary effect of the federal Wages and Hours Act (minimum wage and maximum hour law) as well.

and on consumers are not of particular importance to most of the members and leaders of these unions.

Policies followed will depend in part on the kind of union involved, especially on whether it is a union of highly skilled craftsmen or of a mass of relatively unskilled workers. Unions that can effectively insulate themselves from other groups of workers can pursue successfully policies much narrower in focus than can other unions. A union of highly skilled workers that is able to maintain high initiation fees and apprenticeship requirements may be able to maintain a continuously high-wage policy that would inevitably break down where entrance to a trade was more open. Within a union there will be conflicts of interest where higher wages for some members will mean unemployment for others, and in the weighing of these opposing pressures a leader has very little "rational" basis for arriving at a decision. The winning of a higher wage agreement is a much more obvious accomplishment than is the prevention of unemployment, so that frequently the desire to continue as a union official, or to win leadership from others, results in shortsighted promises and programs to raise wages despite almost immediate curtailments of jobs. Probably union leaders are even more likely to ignore the effects of high-wage policies on the development of new labor-saving devices. It requires labor statesmanship and poise of a higher order to resist the pressures to push up wages.

Policies with regard to hours of work are equally complex. Shortening of hours is regarded as desirable in that it provides more leisure. It is also regarded as a way of spreading work among more members in slack periods. And it is sometimes regarded as a means of reconciling the goals of higher wages and fuller employment. It is probably safe to assume that rarely are union leaders aware of the fact that in reducing hours they may reduce their effective bargaining power for higher wages; only the first and the fourth of the four situations discussed in the preceding section are generally recognized. Again, as in wage policies, the complexity of ends sought and the ignorance of the effects of particular programs combine to make the wise determination of policy exceedingly difficult.

The extent to which members and union leaders are informed of conditions in labor markets, and the extent to which they take into account long-run repercussions of immediate policies will differ markedly from one group to another. Long experience in dealing with employers is likely to lead to more farsighted wage and hour policies, although it may also mean the development of closer supporting restrictions on entry. Policies may be narrow or broad in the extent to which they take into account the interests of workers as a whole. The more the control over union policies shifts from local leaders to national leaders of whole groups of unions, and the more secure unions become in their status as recognized bargaining agencies, the greater the possibilities of real "labor statesmanship" which takes a long view of the effects on labor of the policies of any given time and place. But there is no guarantee whatsoever that these possibilities will be realized.

Finally, policies concerning wages and hours cannot be separated from some of the other purposes most important in the actions of leaders of trade-unions. Frequently demands for concessions in wages and hours are smoke screens or incidental by-products of other programs. When unions or their leaders are primarily concerned with large organizing drives to expand membership, or when they are fighting for a status that will give them future control over labor markets (as for recognition as the bargaining agency, or for closed shops), then the shouting about wages and hours is frequently spurious. Many of these programs would be regarded as very desirable by informed members of the rank and file. In some instances, however, the policy pursued in pushing for wage and hour concessions is in fact a part of the maneuvering of an adroit grafter, perhaps a union official, perhaps employer-paid specialists at inciting trouble. However justifiable or unjustifiable these other purposes may appear to workers and to the public, they complicate still further the problem of explaining and understanding trade-union activities.

Evaluation of Some of the Effects of Policies of Job-Conscious Unionism

In the preceding pages we have analyzed only the most important of the policies pursued by the opportunistic unionism characteristic of American labor. Unions are also interested in many other aspects of control over job opportunities. They set up grievance committees to protect their members from arbitrary actions by employers, especially in the firing of men "without sufficient cause." They take an interest in the conditions of work such as light and air, safety provisions, sanitation, and so on. And in some exceptional cases they operate mutual benefit sickness and burial schemes, or set up cooperative stores for union members. A full discussion of all these and many more activities would require many pages. In evaluating the effects of trade-union actions we shall take some of these "other" programs into account, but we shall make no attempt to cover them all in any systematic way. Our present task is merely to pull together the threads of discussion in this chapter, and to view trade-unionism in relation to the five criteria of social welfare that we have selected for previous discussions.⁵

Trade-unions and the allocation of resources

There are many popular misunderstandings concerning the effects of trade-union activities on consumers. It is obvious that the allocation of resources in accord with consumer preferences is a goal of little concern to trade-unionists seeking to improve their own particular positions in the economy. But it does not follow that everything they do therefore works contrary to the interests of consumers. The incidental effects of their actions on the allocation of resources may be either to further or to interfere with the realization of consumers' preferences. Four principal

⁵ See especially Chapter 9.

groups of policies have the most direct impact on resource allocation: (1) employment exchanges, (2) wage policies, (3) restrictions on entry into unionized occupations, (4) collusion between unions and employers.

1. The services provided by unions in making information concerning job opportunities available to members are sometimes very important. Sometimes unions go so far as to make loans to members to finance shifts to new jobs. But union activities in these spheres have been limited. This is inevitable, since the mobility provided by unions is a mobility only within the job territory covered by the union. To the extent that they do contribute to labor mobility, unions encourage more rapid adjustment in the use of labor resources in response to changing demands. They thus make a contribution toward the realization of consumers' preferences in resource allocation.

2. Wage policies may operate in more than one direction. The allocation of labor resources would be that most consistent with consumers' preferences when workers were hired to the point at which wage equaled the sale value of the marginal product of the workers in each enterprise, and when there are no workers excluded who would be willing to work for less than the existing wage. Such an adjustment tends to take place where there is pure competition throughout. Three kinds of discrepancies are likely to arise between the quantity of labor used in a firm and the preference of consumers with regard to the allocation of labor services: (a) Monopolists restrict the level of production short of the point at which wages equal the sale values of marginal products in order to maintain high prices for what they sell; there is no direct way in which a trade-union can break into such monopolistic controls since they are based on the selling position of the firm and not on its position in the buying of labor. (b) Monopsonists restrict the level of production short of the point at which wages equal marginal revenues in order to push down the prices they pay for productive services; unions may and in fact do exert a check on these monopsonistic practices in labor markets when they bargain for standard rates of pay, and by checking monopsonistic powers remove the incentives of employers to curtail labor inputs. In this respect unions in acting for their own interests act consistently with the interests of consumers. (c) Monopolistic labor groups may push up wages to levels such that employment is curtailed while there are still many workers who would be glad of jobs in the unionized occupation at wages equal to or less than those being paid; this is almost always what happens when trade-unions attempt to raise wages in purely competitive labor markets, and it is a common occurrence when they make excessive wage demands even in monopsonistic labor markets.

3. Restrictions on entry into unionized occupations (and into local labor markets) are obviously interferences with that allocation of resources that would be most completely consistent with the realization of consumers' preferences. Whatever the form these restrictions may take, the result tends to be the employment of fewer workers in the unionized occupations, of more workers elsewhere.

4. Little has been said thus far concerning collusion between employers and trade-unions. This is not a common practice in general, but ever since the end of the nineteenth century it has been one of the many curses of the building trades. Unions get together in rings with contractors and materials dealers. The unions agree that they will not work for "outside" contractors, or materials dealers, and so on, and they get similar guarantees in return. The effects of such agreements are to tie up whole areas of productive activity, keeping newcomers out very effectively and restraining any kind of competition that might lower costs and ultimately increase the outputs of the products involved. Each of these groups supports the exclusive controls of the others.⁶

Realization of workers' preferences

Just as union activities cut two ways so far as contributing to the realization of consumers' preferences are concerned, so also they cut two ways in their effects on the realization of laborers' preferences concerning jobs. These preferences are concerned with which jobs are wanted, with conditions of work on the job, with hours of work, with security of job tenure. We shall consider each of these phases of the problem briefly:

1. For the most part unions discourage rather than encourage mobility between different jobs. This is especially true of the craft unions, though the industrial unions of the C.I.O. are beginning to weaken some of these exclusive controls in certain areas of production. Any restriction set up by a union on mobility into (or out of) an occupation is a restriction on the opportunities of the excluded workers: it is an interference with the

⁶ In discussing labor and antitrust before the American Economic Association in December, 1941, Corwin Edwards cited a number of examples. (*American Economic Review*, Vol. 32 No. 1, Part 2, Supplement, March 1942. "Public Policy toward Restraints of Trade by Labor Unions; An Economic Appraisal," pp. 443-444.) "In building, withdrawal of union labor was the usual way of enforcing the efforts of contractors to fix prices through bid depositories. It was likewise the device by which organized distributors and organized labor tried to prevent the distribution of plumbing equipment through mail-order houses or other channels which would sell it direct to the consumer; by which organized tile manufacturers, contractors, and labor sought to prevent the sale of tile through jobbers; by which produce dealers and truck drivers forced independent dealers to join a trade association; and by which bakers and bakery drivers combined in various cities to prevent or restrict the sale of day-old bread for human consumption. . . . Attorneys for a branch of the bakery sales drivers union and a large chain store called at the Department of Justice a few months ago and jointly declared that the union intended to force the chain to raise the price of its private brand bread by about 40 per cent or else take it off an urban market. The union attorney declared that if the chain continued to charge low prices, other chains would follow suit, independent bakers would do likewise, eventually some independents might go bankrupt, and in that case some bakery drivers might lose their jobs. Various ways of making the union's demand effective without subsequent indictment of the chain store were suggested by both attorneys, and the union attorney said that since the demand for higher prices originated with the union, he was willing to help the chain avoid indictment. As a result of subsequent investigation a federal grand jury has indicted, not the union alone, but three chain stores and local branches of two unions, alleging a joint conspiracy to fix the prices of bread in that locality."

effective realization of the job preferences of these workers. There is little that unions do to facilitate such inter-occupational mobility.

2. The effects of unions on conditions of work are constructive in giving voice to workers' preferences. When an employer hires a worker he pays him a certain wage for certain hours of work *under certain conditions of work*. Lighting, sanitation, protection against accident, and so on, all of these are important aspects of the job. But in the existing system there is only very limited opportunity for workers to express their preferences for some conditions of work over others. Would they be willing to accept lower wages in order to get better work surroundings? how much lower wages? and for what aspects of improvements in surroundings? The difficulty here is analogous to the imperfection of the retail market in registering consumers' willingness to pay higher prices in order to get credit, pleasant surroundings in which to shop, and so on. Here the union may step in as a representative of the workers, and through its bargaining and discussion it may facilitate the attaining of many improvements in working conditions that are very much desired by individual workers but for which they had no opportunity to express their preferences. It may of course go beyond the level for which workers are in fact willing to pay in wages sacrificed, and it may stop short.

3. A similar situation arises in connection with controls over hours of work. Here again the channels for registering workers' preferences are usually very imperfect. They remain imperfect when agreements in this respect are in the hands of the union. Union leaders may manage to bargain for a level that expresses very well the wishes of the majority of workers, or they may seriously err.

4. Some unions, in fact most of the well-established unions, have placed considerable emphasis on the ensuring of job tenure. Obviously the union grievance committee in protecting workers from arbitrary discharge is providing a service for which the members are most willing to pay. When it comes to things like seniority rules and jurisdictional disputes, we must ask whose preferences we are concerned about. How do we weigh one man's preferences as compared with those of another? Conflicting interests within the laboring groups present us with a serious problem here, even though the policy of a particular union may very adequately express the preferences of its members. Many of the union attempts to provide security of tenure for members run counter to preferences of workers in other occupations or outside the union. If every man is to be regarded as having an equal right to determine what shall be done, then it is probably safe to say that unions interfere with the attaining of workers' preferences in that the majority is likely to be on the "outside." This becomes less and less true as unionism spreads to include larger and larger proportions of the population.

Equity in income distribution

In previous chapters we have discussed three ways in which people sometimes look at equity in the distribution of income: (1) distribution according to economic contribution, (2) the provision of at least a

minimum for all, and (3) the modifying of existing inequalities. Do trade-union activities lead us nearer to or further away from each of these various goals?

1. Anything that interferes with the realization of consumers' preferences in the allocation of resources is also an interference with the distribution of income according to the potential contributions of all workers. Thus trade-union policies that restrict entry to certain occupations or that set wages above competitive equilibrium levels are preventing potential workers from making the contributions that they might and from securing the corresponding wage. On the other hand, only under exceptional circumstances can a union force employers to hire workers when the resulting marginal revenues would be less than the wages. The workers remaining employed may receive wages equal to or less than their marginal revenues; they will not ordinarily receive wages greater than their economic contributions in the sale value of marginal products. When employers are monopsonists or oligopsonists, they generally pay wages less than the economic contributions of the employed workers. Here a union may step in and force wages up so that the workers employed will receive incomes more nearly equal to their actual economic contributions. In this respect trade-union actions are consistent with obtaining for workers wages more nearly in accord with the productivity criterion of equity in income distribution.

2. Unions can accomplish little in obtaining a decent minimum wage for those otherwise below what would generally be regarded as a decent living. These groups are generally unskilled and it is very difficult to develop any kind of continuous pragmatic unionism. The very poverty of the membership is itself a source of weakness in union organization since it is difficult to finance union activities. The membership is almost inevitably very shifting. These difficulties are revealed in the history of trade-unionism in this country. The strongest pragmatic unions have been among highly skilled workers. Unions of the mass of workers most frequently turn from opportunistic job-bargaining to more ephemeral ideas of social reform or even social revolution. There is no concrete common interest to hold them together. The development of industrial unionism along relatively practical lines is a step in the direction of protection and bargaining for the unskilled groups, but cases in which they touch the lowest income groups remain the exception.

3. For the same reasons, trade-unionism is a very poor device for modifying income inequalities among workers. The usual effect of most trade-union policies is in fact to increase these inequalities, by excluding "outsiders" from some of the more highly paid occupations. Insofar as union action can bring about a redistribution of income as between workers and propertied groups, however, it can contribute to modifying some aspects of income inequalities. This is possible only in limited ranges of union action, especially where a union is able to get for its members incomes which were previously part of the profits of monopsonists. Any general effect of trade-union actions on the relative returns

to owners of material productive agents and lenders of funds as compared with returns to labor is very doubtful.

Security and stability

The effects of trade-union policies on the degree of security and stability in the economic system as a whole are a matter of some debate. Moreover, the analysis of these effects can be undertaken only after we have examined business fluctuations. Within the particular job territory covered by a union there are, however, some relatively obvious effects of trade-union policies. Grievance committees give greater security to individual members, so also do seniority rules. Restrictions on outputs are frequently intended as a protection against immediately pending unemployment, and may have that effect in certain limited kinds of situations. On the other hand, restrictive practices of trade-unions may threaten the security of other workers; and mistaken high-wage policies may lead directly to technological displacements and technological unemployment.

Raising planes of living

Trade-union policies of various kinds affect directly the size of the national real income. We shall consider some of them briefly under the following classifications: (1) policies that increase the efficiency of the workers, (2) policies that restrict the amount of work done per employed worker, (3) policies that affect the amount of employment per man per lifetime, and (4) policies that check the introduction of new lower-cost techniques.

1. Shorter hours, better surroundings in work, and sometimes higher wages, seniority rules, output restrictions and participation of labor in management, may all contribute to greater efficiency of the workers involved. The effects of shorter hours and better work surroundings on both physical efficiency and morale are obvious. When wages have been exceedingly low, an increase may have similar effects. Output restrictions are sometimes introduced to protect workers in piece-rate establishments from ruthless pressures to work at a rate that wears them out and ultimately leads to diminished efficiency; in such cases immediate restrictions sometimes mean ultimately greater product. Seniority rules may improve morale and stimulate workers to at least an average level of performance, so that they will not be fired and lose their seniority ratings. On the other hand, such protection removes some of the incentives of competition from younger men. Labor participation in management is most important in its effects on morale, as a means of bringing about greater labor-management cooperation. It may also be a means through which employers can learn of ways of economizing that appear to workers or the foreman on the job and would not otherwise be passed on to management.

2. Some of the policies that may increase labor efficiency may be carried beyond this point and result in diminished productivity per worker. The effects of reducing hours and restricting output in piece-rate establish-

ments are sometimes of this sort. Also some actions of unions are clearly restrictive of production.

Most obvious of these is the use of the strike as a weapon in labor bargaining. During a strike, production ceases completely. There is obvious waste in the idleness of labor resources and of some of the other productive resources employed in the closed firms. There may also be indirect repercussions of strikes on production of goods sold to the firms suffering from strikes and of goods made from the products of such firms. Many of the economies of efficient planning may be lost. These losses seem more serious in those situations in which the strike seems to serve no otherwise justifiable purpose. It is for this reason that jurisdictional strikes have been so often decried by labor leaders and employers alike.

Losses resulting from sabotage are as obvious as those resulting from strikes. Such actions are probably exceptional, but where they occur they are very disrupting to productive activity. Sometimes the sabotage is very evident, sometimes it is more subtle and difficult to get at. Rarely are such practices approved by leaders of well-established unions.

Important especially among the skilled workers in building trades and the theatre are some of the attempts to spread work among members by restricting the outputs of each man. One of the best known is the long-standing rule of the bricklayers unions in many localities that no member may lay more than a specified number of bricks per hour. For an individual union at a particular time and under certain circumstances, this may appear as a very beneficial procedure. It may actually succeed in stretching out the work. But decreased marginal products are very likely to lead to decreased employment. Unfortunately when a whole group of unions whose members contribute toward the production of a joint product such as housing follow such practices, no single union can, by changing its policy, have enough effect on the cost of the jointly produced houses to make much difference in the number of houses that will be constructed. Hence the demands for their services may be quite inelastic, and restrictions on outputs may work remarkably well. If all the unions involved could get together to eliminate their restrictive practices, all might together benefit. No one acting alone is able to do so. For the entire group of unions it makes no sense at all to assume that there is just so much work to be done and that it can be spread out by restricting outputs. The removal of restrictions that would be advantageous to consumers would also benefit the large group of unions; but when each union acts separately it is likely to act inconsistently with these wider interests.

Finally, productivity per man is reduced by actions that remove incentives to greater efficiency. This is especially likely to follow from some of the policies of restriction of entry practiced by a few of the stronger unions in the skilled occupations.

3. Union policies affect the amount of work per man per lifetime in a number of ways. We have already pointed out that we cannot at this stage evaluate the effects of trade-union policy on business fluctuations

and hence on depression unemployment. There are two other important ways in which trade-union actions affect the working life of laborers. First, insofar as they improve health through their controls over hours and working conditions, unions prolong the potential working lives of their members and others who benefit from their bargaining. Second, seniority rules, in the few cases in which they exist, frequently prolong the working lives of the union members.

4. Unions fight the introduction of new lower cost techniques that threaten to displace them in all or a part of their job territories. Various regulations of a restrictive sort have been developed in the attempt to protect these vested interests. In a now famous case, the Chicago painters and glaziers union refused to install cabinets that had been painted and had had the glass installed at the factory, insisting instead that the paint be scraped off and the glass taken out, so that the cabinets could be painted, assembled, and installed according to union rules. For years the stonecutters tried to outlaw the planing machine, and refused to work with machine-cut stone. Following the introduction of "talkies" the musicians' union insisted that theatre orchestras be maintained at full pay even though no longer used. In a few cases unions obtain desired restrictive ends through indirect means such as the use of building codes to protect craftsmen against the introduction of new building materials and prefabricated housing. The line between some of these practices and clear "labor racketeering" is a very thin one; and many of the worst examples are in fact the result of gangster control in a few of the unions, especially in building trades in some of the biggest cities. In their most extreme form, such practices are much decried by most union officials, who see in them a threat to the good name of trade-unionism everywhere, no matter how statesmanlike its policies may be. Moreover, there are many union leaders who cooperate with employers to facilitate gradual adjustments to technological change.

CHAPTER 33

Government and Trade-Unions

TO ATTEMPT to discuss the position of trade-union activities in the law and the role of government in dealing with issues in which organized labor is involved is to tackle a very elusive set of problems. The position of labor organizations both in fact and in the law is never long the same; conditions and attitudes are continuously shifting. Moreover, currently we seem to be entering upon a new era in the formulation of public policy toward labor. Many kinds of policies are recommended and there are many genuine differences in opinion as to what policies would be most effective in protecting all interests concerned and furthering the general welfare.

Discussion divides naturally into two major parts, the power conflict between workers and employers and the interest of the general public in labor practices which restrain trade. The first section of this chapter will be devoted to an examination of the power positions of the two primary parties to labor bargains. Then, after a brief discussion of techniques in the peaceful settlement of disputes, we shall turn to the questions of public policy and labor activities in restraint of trade.

The Legal Powers of Labor and Employers

Traditional principles in the law

Government may intervene in labor conflicts by legislating, by executive action, and by judicial decisions. Until recently the most important continuous avenue of approach was clearly the judiciary, and this is probably still the case despite the increasing role of legislation and intervention of executive agencies. Much of this judicial action is in fact the interpretation and application of legislative action.

There are two major precepts that underlie most of the law and of judicial decisions in cases pertaining to organized labor prior to the 'thirties, and that still pervade much of the thinking on these matters. The first rests on the individualistic interpretation of social processes that emphasizes the importance of *freedom of contract*; this involves an assumption of equal bargaining power for both (or all) parties to a contract. The second rests on a concept of the sacredness of *property rights* very broadly interpreted to include the right to carry on business with a

minimum of interference with entrepreneurial decisions. The implications of these two concepts in labor cases deserves further exploration.

1. Freedom of contract. The doctrine of freedom of contract in the labor bargain means freedom of the worker to choose where and for whom and under what conditions he will work, to accept or reject a job contract as he may see fit. It means freedom for the entrepreneur to decide whom he will hire and under what conditions, to accept or reject the terms offered by each workman. There is to be no coercion on either side; the appearance of such coercion invalidates the contract. If either party to the contract breaks the terms the other is released from it and could bring suit. If an employee fails to do his work he may be fired at any time, and if an employer fails to pay the promised wages or to provide the promised conditions of work the employee may quit at any time. Each may sue the other.

This doctrine of freedom of contract is clearly based on a fundamental assumption that the two parties to the bargain are in comparable positions; that the freedom of the worker to quit, for example, is the equivalent of the freedom of an employer to fire a man; that the freedom of a worker to accept or reject the terms of a job offer is equivalent to the freedom of an employer to accept or reject a man seeking employment. This equality never fully existed; but with the growth of large business enterprise its inapplicability is glaringly obvious. The corporation as a "legal person" is clearly in a very different power position from the individual workman seeking an opportunity to earn a livelihood. The individual worker's chief protection is the competition of rival employers for labor. Wherever there is monopsony or oligopsony in the buying of labor services this protection disappears. A recognition of this inequality in fact has led to progressive modifications of the attitudes of courts toward labor organizations as bargaining agencies representing the workers. Finally, a recognition of the unequal bargaining positions of corporations versus individual workers has found expression in recent labor legislation,¹ which abrogates the earlier extreme interpretations of the doctrine of freedom of contract in labor bargains. Focus is shifted to the problem of balancing the power of employers and employees in bargaining in the labor market. Freedom of individuals from coercion in labor bargains is reinterpreted in relation to developing conditions.

2. Property rights. Property rights have been very broadly interpreted in the court history of the United States; they include the right to do business in such a way as to maximize entrepreneurial income within the framework of the law. Legal restraints on business activities have been regarded as necessary evils to be introduced only to the extent that they are necessary to protect the general public and the freedom of others. Individuals and private groups have no "right" to interfere with these rights of businessmen to carry on their activities unrestrained.

¹ NIRA. Section 7 a. and the National Labor Relations Act; see page 491.

It must be immediately evident to anyone who gives thought to these questions that there will be continuous conflicts concerning what acts of what groups constitute justifiable expression of the rights of these groups and what acts constitute unjustifiable interferences with the rights of others. For example, in labor conflicts the right of the entrepreneur to conduct his business without coercive interference would appear to outlaw strikes; on the other hand, unless they have signed contracts for employment for a certain length of time individual workers have a right to leave their jobs at any time. In conflicts between these sets of rights the courts are called upon to make decisions.

Despite the broad interpretation of property rights in protecting businessmen, there is no equivalent doctrine concerning rights of workers. Labor leaders have continuously argued that the "property right of the worker in his job" is just as fundamental a right as is the property right of the employer in his business. In time of strike, the workers on strike regard the job as still belonging to them. Picketing is a technique for maintaining their control over their "job rights." Until recently this interpretation has had no backing in the courts. It comes directly into conflict with the employer's interpretation of his rights to hire and fire as he may choose. Clashes over issues of this kind have been repeated in American history.

The right to bargain collectively

Pragmatic unionism in the United States functions primarily through collective bargaining. The claim to the right to bargain collectively has gradually gained increasing acceptance with the growth of strong employer groups on the one hand and effective labor organization on the other. The large enterprise is seen to have an advantage over employees acting individually, so that some modification of the individualistic interpretation of "freedom of contract" is in order. At the same time labor has gained greater strength in pushing its cause. Three sets of legal rights are basic here: (1) the right of unions to exist, (2) the legal status of collective agreements, and (3) the legal rights to force collective bargaining processes on reluctant employees.

1. **The right of unions to exist.** In the very early stages of trade-union development in England, all unions were regarded as "criminal conspiracies" and those persons involved were prosecuted on these grounds. In the United States the view taken in early cases was less extreme; but since almost all the *acts* of labor organizations were regarded as unlawful under the common law concept of "criminal conspiracy" the effect was much the same. The right of unions to exist was pretty well settled, however, by a case appearing in the Supreme Court of Massachusetts in 1842 (though this has not been held to apply to radical revolutionary unions). The Court declared that nonprofit associations of workers were not *as such* criminal conspiracies. This meant that so long as a union went about its own business much as a fraternal society it was to be left alone.

2. Collective agreements as legal contracts. As soon as a union tried to do anything beyond a mutual benefit program, as soon as it came in contact with outside groups, this early policy of nonintervention ceased. Such aggressive activities of unions were regarded as a matter of greater public concern; there was inevitably interference with the freedom of employers to run their businesses in their own way. On the other hand, collective bargains as such are peaceable instruments for the adjustment of economic relations. Thus while many of the techniques that established the power of unions to drive effective bargains continued to be subject to attack in the courts, the agreements themselves have in general been looked upon with favor. Pleas brought by workers claiming that employers were not living up to the terms of collective agreements have been sustained by the courts. Usually damages have not been allowed, but employers have been ordered by the court to live up to the terms of the contract.

3. Guarantee of the right to bargain collectively. Increasingly favorable attitudes toward collective bargaining have culminated in recent years in the National Labor Relations Act of 1935. This act grew directly out of experience in applying Section 7-a of the National Industrial Recovery Act.² The provisions of that section were:

Every code of fair competition, agreement, and license approved, prescribed, or issued under this title shall contain the following conditions: (1) that employees shall have the right to organize and bargain collectively through representatives of their own choosing; and shall be free from the interference, restraint, or coercion of employers of labor, or their agents, in the designation of such representatives or in self-organization or in other concerted activities for the purpose of collective bargaining or other mutual aid or protection; (2) that no employee and no one seeking employment shall be required as a condition of employment to join any company union or to refrain from joining, organizing, or assisting a labor organization of his own choosing; and (3) that employers shall comply with the maximum hours of labor, minimum rates of pay, and other conditions of employment, approved or prescribed by the President.

To the workers this appeared to give protection against many of the weapons previously used by employers to break the unions, especially company unions, yellow dog contracts, and discharges for union activities. The employers interpreted the act quite differently and a tremendous surge of organization of company unions ensued. Conflicts were almost continuous.

With the *Schechter* case,³ which declared the NIRA unconstitutional, Section 7-a was, of course, automatically dropped; but the new National Labor Relations Act was immediately passed to take its place. Those who formulated this new law had learned something from the experiences under the NIRA and tightened the provisions to leave no doubt of the intent of the act to guarantee genuine representation of workers in collective bargaining. The federal power in the National Labor Relations Act rests on the power of Congress to regulate interstate commerce.

² For a general discussion of this act, see Chapter 26.

³ *Schechter Poultry Corporation vs. United States*, 295 U. S. 495 (1935).

The agency created to administer the act is the National Labor Relations Board. *Its basic function is judicial; it does not act as a mediator or arbitrator in settling the terms of employer-labor agreements.* It protects workers against unfair labor practices of employers that would interfere with the *process* of collective bargaining, not with the *terms* of the bargain. In cases of dispute over which is the appropriate organization to represent workers in bargaining, the Board makes a decision. In many cases this decision is reached by holding elections to determine employee's choices. When employers continue with unfair labor practices after the Board has issued an order to "cease and desist," the Board may turn to the Circuit Court of Appeals for an injunction to enforce the order. The employer may also appeal against the order.

The National Labor Relations Act gives a detailed statement of the case for collective bargaining. It says that individual labor bargaining leads to depressed wages and hence lowered purchasing power and depressed general economic conditions; this is a highly questionable argument, which would be regarded as unsound by many economists. It also states that inequalities of bargaining power prevent equalization of wage rates between different industries; this argument gains strength from the fact that those labor markets in which unions need the protection of the law in order to negotiate collective agreements are in fact the areas where workers are weakest in dealing with big industrial magnates (especially steel and automobiles). The Act holds that experience proves that protection of the right to organize and bargain collectively insures much greater moderation in labor strife than would otherwise prevail, and that such bargaining provides a medium through which a much larger proportion of conflicts can be settled peaceably and equitably. Thus there is in this recent development of labor policy a very strong positive support for collective bargaining, and a very significant modification of earlier individualistic interpretations of freedom of contract.

The law and the weapons of labor

The three principal weapons of labor are the strike, picketing, and the boycott. All of these are inevitably in some measure coercive, and they have therefore come under attack on many occasions.

1. The right to strike. The right to strike for just cause has generally been recognized by the courts in recent generations. There is very little legislation on this matter. Waves of public protest against strikes have led from time to time to recommendations for such legislation; but they have not resulted in action. In the courts, strikes for higher wages, shorter hours, safety devices, or other direct benefits to the union members have generally been regarded as legitimate. Strikes for union recognition, strikes against competing unions, strikes for closed shops, and sympathetic strikes have frequently been regarded as illegal in the past.

Strikes for union recognition and against rival unions have taken on new significance in recent years. With the National Labor Relations Act, workers are protected in the right to organize, so that we might expect these strikes to be eliminated, but this is not the case. This legislation coincided with a period in which labor organization was expanding on a tremendous scale, and it lent support to such expansion. Strikes for recognition in the mass-production industries brought the case of the workers to the attention of the National Labor Board, so that for a time such strikes continued. In view of the provisions of the Act these strikes had a new legal support that contrasted with the earlier situation. After this big organizational drive was once fully launched and collective bargaining of independent (noncompany) unions established, there was of course a falling off in such strikes. But they reappeared in another form—as conflicts between rival unions competing to represent the workers in collective bargaining. These issues are complicated by jurisdictional strikes in which two unions are in conflict over which group of workers should do a particular job (as hanging metal doors). In recent cases the United States Supreme Court has declined to take action under the antitrust laws⁴ against these jurisdictional strikes.

Strikes for the closed shop have generally been opposed in the courts on the grounds that closed-shop agreements constitute interferences with individual freedom; they interfere with freedom of the employer to hire and fire whom he may choose; they interfere with freedom of the individual worker to work wherever he can find a job regardless of his membership in a union. The issue has been very much debated recently under the National Labor Relations Act, but it has not arisen in court.⁵

"Sympathetic" strikes are strikes of a union not directly involved in a dispute, in sympathy with another union that is striking. These strikes are very common in closely related crafts, in the building industry, for example. Or textile workers in firm B may strike against "struck materials" when the employer of firm A where workers are on strike attempts to transfer the filling of his orders to firm B. Such strikes are analogous to log-rolling in Congress, when the representative from Nevada and the representative from Massachusetts trade votes on a silver purchase program and a tariff on woolen goods. Sympathetic strikes may become very widespread, in extreme cases reaching the dimensions of the "general strike" in Great Britain in 1926. When they reach such a scope sympathetic strikes are no longer primarily pragmatic; they verge rather on mild class warfare through an expression of class solidarity. State courts have taken differing attitudes on sympathetic strikes; the most important cases in the United States Supreme Court have ruled against them.

⁴ See page 489.

⁵ Labor interpreted Section 7-a, point (2) as protection against yellow dog contracts and company unions. Employers interpreted it as a ban on the closed shop. The same issue arises under the NLRA.

2. The right to picket. There have probably been more court cases concerned with picketing than with any other single phase of union activity. This is to be expected, since picketing is more likely than any other action to be associated with violence and public disturbance. It involves conflicts of interest directly of concern to striking workers, nonstriking workers and "scabs," employers, and general public. Whoever may be at fault where physical violence appears, blame is likely to be focused on the union. Fundamental issues over individual freedom and property rights are immediately raised. In general courts have supported picketing activities that could be clearly regarded as "peaceful persuasion"; and they have declared against "intimidation." But almost anything can be and has been regarded as "intimidation." Yet despite the generally unfavorable attitudes of courts, unions have continued to picket in order to protect the striking worker's "right to his job"; they have frequently made genuine efforts to keep that picketing peaceful and orderly. Employers have a clear incentive to incite disorder so that they may be in a strong position in the courts and in the public's attitudes.

3. The right to boycott. A primary boycott in which workers refuse to buy the goods produced by their employer is generally regarded as legal; but this limited type of boycott can rarely exert much pressure on the employer. Any attempt to organize a "secondary" boycott by the customers of a firm stands on very shaky grounds in the courts. This is especially true since such cases have been attacked as restraints of trade under the antitrust laws. The most famous case, of the Danbury Hatters, was in 1903, when triple damages were levied against the union (totaling three times the estimated damage of \$80,000 to the employer's business). In another case, after the Clayton Act in 1914, the Court took the same position; and so the "secondary boycott" stands outlawed in the precedent of Supreme Court decisions today. On the other hand, more recent state cases are not so clearly in accord with this position; and there have been no recent attempts to apply federal antitrust laws in boycott cases.

The law and the weapons of employers

The powers of employers in labor bargaining have gradually come under more restrictions while labor has been gaining more recognition in courts and law. Employers' associations for collective action have been upheld as essential to the protection of property rights.

The lockout has never been questioned except as it may be associated with breach of a union-employer contract; but where such breach of contract is involved the employer may be ordered by the Court to refrain from breaking the agreement, and damages may be claimed. The first case on this point was in 1937. On the other hand the employer's right to keep his plant open during strikes and even to have protection of government police or militia in keeping his plant open has rarely been questioned. Recently there have been some attempts to limit or prevent the transportation of strikebreakers from other states.

The National Labor Relations Act has significantly limited the powers of employers by outlawing yellow dog contracts and by requiring bargaining with representatives selected by the workers themselves without entrepreneurial intervention (hence breaking company unions as employer weapons). It also prevents discrimination against union men in either refusing to hire them or in cases of discharge.

The strongest of all employer weapons against unions has been the use of *injunctions*. Injunctions are usually issued in time of strikes. We may repeat briefly what they are: They are court orders restraining the workers and their leaders from certain acts. If such acts are performed after issue of an injunction the offenders are guilty of contempt of court. The principal grounds on which such injunctions are granted to the employer or group of employers requesting them is the protection of property, interpreted broadly to include the protection of business income.

Labor has fought this weapon for years, but until recently with very little success. Provisions introduced in the Clayton Act in 1914 were regarded by labor leaders of that time as protection against unrestricted use of injunction proceedings⁶ but this proved an illusion. Finally, in 1932, the Norris-LaGuardia Act was passed. Provisions are detailed and explicit, and corrections of procedure give defendants in contempt proceedings much more legal protection than was previously available.⁷

⁶ *Senn vs. Tile Layers' Protective Union*, 301 U. S. 468 (1937).

⁷ Section 4 of this Act specifies restrictions on the use of injunctions as follows: Sec. 4. Restriction on injunctions.—No court of the United States shall have jurisdiction to issue any restraining order or temporary or permanent injunction in any case involving or growing out of any labor dispute to prohibit any person or persons participating or interested in such dispute. (as these terms are herein defined) from doing, whether singly or in concert, any of the following acts:

(a) Ceasing or refusing to perform any work or to remain in any relation of employment;

(b) Becoming or remaining a member of any labor organization or of any employer organization, regardless of any such undertaking or promise as is described in section 3 of this act;

(c) Paying or giving to, or withholding from, any person participating or interested in such labor dispute, any strike or unemployment benefits or insurance, or other moneys or things of value;

(d) By all lawful means aiding any person participating or interested in any labor dispute who is being proceeded against in, or is prosecuting, any action or suit in any court of the United States or of any State;

(e) Giving publicity to the existence of, or the facts involved in, any labor dispute, whether by advertising, speaking, patrolling, or by any other method not involving fraud or violence;

(f) Assembling peaceably to act or to organize to act in promotion of their interests in a labor dispute;

(g) Advising or notifying any person of any intention to do any of the acts heretofore specified;

(h) Agreeing with other persons to do or not to do any of the acts heretofore specified; and

(i) Advising, urging, or otherwise causing or inducing without fraud or violence the acts heretofore specified, regardless of any such undertaking or promise as is described in section 3 of this act.

The Adjustment of Labor Disputes

There is a third party involved in every labor dispute. The community at large has a great stake in the maintenance of orderly labor relations. Hence we must concern ourselves with the problem of appropriate machinery for adjusting labor disputes and maintaining industrial peace by preventing their occurrence.

Once trouble has broken out, *conciliation* is the least drastic instrument and generally the one invoked first. Mayors, governors, judges, bishops often appeal to the parties, offering facilities for them to come together and settle their grievances peaceably around the conference table. Any party, public or private, may seek to bring the disputants together, but usually a good deal of prestige is required, by way of personal reputation for integrity or prestige of office. If this outsider takes any part in the negotiations beyond bringing the parties together, the process is known as *mediation*. Usually a mediator is a go-between and confidant of both parties, who ascertains by separate conferences what sort of compromise is likely to succeed before bringing the parties face to face to attempt a final settlement. Conciliation or mediation can only hope to be successful if the workers are a closely organized and disciplined group so that their leaders are capable of entering into an acceptable settlement. The Conciliation Service of the United States Department of Labor has a fine record in adjusting disputes through mediation and stands ready to offer its services on request.

If mediation is a diplomatic function, *arbitration* is a judicial one. In arbitration an impartial third party hears the claims of the disputing parties and renders his decision as to the merits of the case and the appropriate solution. The parties may or may not agree in advance to accept his award as a basis of settlement. Referring the dispute to arbitration in the first place may be either compulsory or voluntary. Voluntary schemes seem to offer more prospects of success, since the parties are less likely to feel that they have been coerced.

The machinery of arbitration may be either continuous or "ad hoc," created specifically for every new occasion. The continuous machinery may on the one hand consist of a permanent body to which disputes may be voluntarily referred. Great Britain has a system of Industrial Courts of this type. The disputes referred for arbitration may be serious, such as strikes or threatened strikes, or they may be minor adjustments not serious in themselves, but which could rankle and have serious consequences if left unsettled, to add their force to that of other minor grievances. The permanent arbitration machinery might on the other hand consist of a body set up for compulsory arbitration of disputes, such as the former short-lived Kansas Court of Industrial Relations, or the New Zealand Arbitration Court. Ad hoc arbitration of course has to be voluntary in character, set up in view of a specific crisis.

A very interesting approach to the problem is exemplified by the Canadian Industrial Disputes Investigation Act. This act applies only

to a limited number of industries, but requires that no employer shall change the conditions of employment, and no union shall strike, without giving thirty days notice of intention to do so. During this time the Minister of Labor might appoint a board of conciliation and investigation, consisting of one representative from each side and a neutral chairman. This board should attempt to bring the parties together in a voluntary settlement, but failing this, might proceed to investigate the facts of the situation and make a report recommending a settlement. After the findings are published the parties may accept or reject the proposals as they see fit, and are free to strike or change conditions of employment. The delay is supposed to give the parties time to "think things over," and to inform public opinion on the merits of the case and thus contribute to a just solution. This act has been very successful, as much because of wise and tactful administration as on account of the merits of the act itself.

In appraising ways of dealing with actual or potential industrial disputes two general principles seem to be important. First, it is desirable that the machinery should be continuously available in order to prevent a number of minor grievances from "snowballing" into a mighty discontent. Crisis arbitration is at a disadvantage right from the start. Second, the machinery should leave as much room as possible for voluntary action, as the psychological effect is much more salutary and conducive to more lasting results.

Perhaps no machinery can be quite so conspicuously successful as continuous collective bargaining between a union and an employer who respect and trust each other. The history of the Amalgamated Clothing Workers Union in its relations with employers has been very instructive in this regard. This union has had exceptionally brilliant and responsible leadership, so that it might be too much to expect a letter-perfect repetition of this success in other industries.

Trade-Unionism and Restraint of Trade

Ever since the passage of the Sherman Antitrust Act in 1890, the anti-trust laws have been used by employers to fight trade-unions. Reference has already been made to the use of this law as a basis for suits against unions engaging in secondary boycotts and as an added basis for injunction proceedings. In these cases antitrust action was initiated primarily at the request of employers as a weapon against labor. Recently there has been a new emphasis in the application of antitrust laws to labor. The increased activity of the antitrust division in its efforts to break up restraints in the American economy in the interests of the general public has included suits brought by government against certain types of trade-union policies as well as against business and agriculture.⁸ It is this approach to which we now turn our attention.

⁸ Mr. Arnold explains the attitude of some labor leaders who oppose the present policy of the Department of Justice in the following way:

"Labor has been led to believe that the Sherman Act is a weapon given to their

Current opinion is split into two major factions. One of these is headed by Thurman Arnold and his followers in the antitrust division of the Department of Justice. The other group, primarily led by representatives of skilled labor, attacks Mr. Arnold. We shall divide our discussion into three major parts: (1) current court decisions in antitrust labor cases, (2) Mr. Arnold's arguments, and (3) arguments of the opposition.

Recent Court decisions in antitrust labor cases

There have been two important cases that have recently come up in the Supreme Court of the United States in which applications of the anti-trust laws to labor cases were involved. The first of these was the *Apex* case in 1940, the second the *Hutcheson* case in 1941.

In the *Apex Hosiery* case⁹ the mill owners brought suit against the American Federation of Full-Fashioned Hosiery Workers for damages to property and loss of business during a sit-down strike. The case was brought under the Sherman Antitrust Act, on the grounds that this interference with production constituted restraint of trade. The majority opinion of the Court was delivered by Mr. Justice Stone. It held that the Apex Company could not recover damages in this case; but at the same time the opinion stated definitely that the Sherman Act was applicable to labor cases where there was clearly *monopolistic intent*. The statute was regarded to be inapplicable to labor unless "the court was of the opinion that there was some form of restraint upon commercial competition." It declared that the restriction on production and hence on inter-

enemies for the purpose of harassing them in their legitimate objectives. Nothing can be further from the truth than this belief. The Sherman Act condemns no reasonable use of the collective power of labor unions. Properly construed the Act is a charter of freedom for labor organizations—a protection against the aggressive tactics of others who seek to hamper them in the legitimate use of their power. Some labor leaders know this, but many are confused.

"What is the reason for this confusion? In my opinion the reason lies in twenty-five years of *private* prosecution of labor under the Sherman Act by organizations charged with no public responsibility many of whom have had the destruction of labor unions as their principal aim. . . . Private corporations used the Act to police strikes a few years ago through the weapon of injunction, and today through the weapon of suits for treble damage. . . . The total damages sought against labor unions today in pending, private suits under the antitrust laws runs over \$10,000,000. These suits have the potential power of crippling the entire labor movement. Even if labor wins the cost of defense is staggering. They have given labor the impression that the Sherman Act is a private instrument designed solely for its opponents. . . .

"Enforcement of law in private hands in the nature of things cannot be fair enforcement. . . . The remedy of the evils of private enforcement does not lie in abolishing enforcement because no individuals or organizations in business, in labor, or in agriculture can be left free to determine for themselves how far they may go in imposing their power on others. The only remedy for the evils of private policing is to substitute policing by public officers not responsible to private interests." (Arnold, Thurman W., *The Antitrust Laws and Labor*, an address before the American Labor Club, January 27, 1940. Mimeographed.)

⁹ *Apex Hosiery Co. vs. Leader*, 310 U. S. 469 at 487 (1940).

state shipments resulting from the sit-down strike did not "operate to restrain commercial competition in any substantial way," that the union at the Apex plant was directing its efforts only toward the legitimate ends of labor bargaining.

By many people the statement of the Court in the "Apex decision" was regarded as a victory for Thurman Arnold in his attempt to invoke the antitrust laws against labor practices that restrained trade. For example, a writer in *Business Week* in June 1940¹⁰ remarked that despite the decision clearing the union the Court "went out of its way to make a distinction between the facts in the case before it and the general assumption of union liability under the antitrust laws" . . . and that "by holding that unions are subject to the antitrust laws when their actions are directed toward the control of prices or toward the placing of restraints on interstate commerce, it scored a bull's eye for Arnold." Labor groups took a different view of the meaning of the decision; and *The Nation* came out with an article entitled "The Apex Indecision."

With the *Hutcheson* case in October 1941 many of Mr. Arnold's hopes were dashed to the ground. The background of this case may be given briefly. The carpenters and machinists both had agreements with Anheuser-Busch, a company manufacturing beer. The two unions were in dispute as to which should control the installation of metal machinery, and the American Federation of Labor had awarded the decision in this dispute to the machinists. The carpenters had at one time accepted this decision but they later demanded that they be given the job. When Busch refused they went on strike, and organized a nation-wide boycott of Busch's beer. When the case against the carpenters, as restraining trade, was brought to court the Court sustained a demurrer which seemed to grant considerable immunity to unions under the antitrust laws; it said:

So long as a union acts in its self-interest and does not combine with non-labor groups, the licit and the illicit . . . are not to be distinguished by any judgment regarding the wisdom or unwisdom, the rightness or wrongness, the selfishness or unselfishness of the end of which the particular union activities are the means.¹¹

Commenting on their decision, Mr. Edwards of the Antitrust Division states: "In the literal meaning of this language the question whether the antitrust laws have been violated depends not upon what was done in the market but upon what group did it."¹² Three other cases were brought in court to check this interpretation and all of them were discussed by the court without hearings, apparently on the same grounds, that there was no combination of labor with nonlabor groups.

¹⁰ *Business Week*, June 1, 1940, p. 16.

¹¹ *U. S. vs. Hutcheson*, 312 U. S. 219 (1941).

¹² Edwards, Corwin, "Public Policy Toward Restraints of Trade by Labor Unions: An Economic Appraisal," *American Economic Review*, Vol. 32, No. 1, Part 2, Supplement, March 1942, p. 432.

The position taken by the Antitrust Division

The Antitrust Division of the Department of Justice considers that there are important aspects of trade-union action which restrain trade, thereby depressing employment and lowering American planes of living. They argue that these actions should be subject to prosecution and that insofar as they are not regarded as covered by the present antitrust laws there is a need for new legislation.

In taking this position Mr. Arnold and his followers point out that recent legislation, especially the National Labor Relations Act and national wage and hour legislation, guarantees to labor adequate power in attaining reasonable conditions of work, that there is now "less need than before to jeopardize other legitimate interests in order to protect labor interests."¹³ They would not suggest applying antitrust laws against any of the "ordinary" activities of trade-unions such as efforts to obtain union recognition, wage and hour policies, and policies relating to conditions of work "as safety devices and protection against speed-ups." They would also consider the problem of the closed shop as outside the appropriate function of antitrust legislation and protection. They would apply the antitrust laws to *unreasonable* activities, that is to activities in restraint of trade that "have no reasonable relation" to the policies just mentioned. Mr. Edwards of the Antitrust Division grouped their "unreasonable" activities in five categories:

1. *First are "restraints of trade designed to destroy one bona fide union or to transfer work from its jurisdiction to that of another."*¹⁴ This was exemplified in the Hutcheson case. There was no way in which the company could comply with the demands of the carpenters without breaking its contract with the machinists. Two of the cases brought up since and dismissed by the Supreme Court were similar; in each of these, A. F. of L. unions were fighting C.I.O. unions that had been designated as the exclusive bargaining agents by the National Labor Relations Board after election by the workers.¹⁵ In the opinion of Mr. Arnold and those associated with him, the lack of power to prosecute unions for such ac-

¹³ *Ibid.*, p. 433.

¹⁴ *Ibid.*, p. 435.

¹⁵ Commenting on such jurisdictional disputes in an address to the American Economic Association in December 1941, Mr. Edwards stated:

"Conduct of jurisdictional disputes by private warfare victimizes the employer and the workers immediately involved, weakens the union movement, and hurts the general public. Obviously the employer who is deprived of the right to do business because he deals with one union rather than another cannot obtain peace by yielding; for he then will be attacked by the other union. He cannot even obtain protection from the government in his efforts to comply with the government's own order to bargain with a designated union. If he is small and financially weak he may be offered the alternatives of illegal action or bankruptcy. It is equally obvious that the workmen involved may be forcibly deprived of their right to be represented by the union of their choice if the union they have rejected is strong enough to coerce the employer. Although the employer is forbidden to impair this freedom, another union may do so with impunity." *Ibid.* pp. 435-436.

tivities has resulted in a rapid spread and exaggeration of major jurisdictional difficulties.

The function of antitrust laws in connection with such disputes would be the *enforcing of compliance with awards* when a union loses a jurisdictional decision. They could *not* determine which union should have jurisdiction in the first place; such determination rests with labor groups themselves and with the National Labor Relations Board.¹⁶

2. A second type of "unreasonable" restraint of trade is the *prevention by organized pressure "of industrial progress by preventing the introduction of new processes, improved machinery, and new materials."*¹⁷

3. Third is *"the erection of private trade barriers designed to reserve the markets for local producers and local labor."*¹⁷

4. Fourth is *"the requirement that useless and unnecessary labor be hired."*¹⁷

These three types of practices are closely interrelated, as Mr. Edwards points out:

Such restraints usually overlap. Exclusion of new processes ordinarily forces an industry to employ unneeded men. In construction, such restraints often prohibit production in a workshop rather than on the job site, and thus guarantee that the productive activity shall be local as well as handicraft. The requirement that men be paid whether they are used or not often prevents the introduction of new equipment. Exclusion of nonlocal products prevents use of factory methods which require a wide market. Stopping progress, localizing the market, and forcing the employment of unnecessary labor are frequently the joint results of a single union activity.¹⁷

In practice some difficulties of interpretation inevitably arise. It would have been reasonable for telephone operators, for example, to ask for a gradual introduction of dial phones so that adjustments of employment could be made smoothly. It would be reasonable for unions to protect their workers against speed-ups on assembly lines and in the other shops where pressures on workers are such as to impair health and shorten efficient working lives. But such marginal issues do not present insurmountable difficulties in the defining of unreasonable labor practices in restraint of trade.

5. A fifth type of "unreasonable" restraint of trade is *"interference with competition among employers by fixing prices, allocating markets, controlling channels of distribution, forcing enterprises out of business regardless of their labor record, or otherwise directly limiting commercial competition."*¹⁸

¹⁶ Mr. Edwards comments further:

"If boycotts and other restraints of trade designed to upset an award are regarded as unreasonable, the antitrust laws should restrict labor leaders to the processes of negotiation, persuasion, and arbitration in conducting jurisdictional disputes with other labor leaders.

"Legal protection for unions against forceful destruction by other unions would also go far toward a solution of some problems of union democracy which are hard to deal with directly." *Ibid.*, p. 438.

¹⁷ *Ibid.*, p. 439.

¹⁸ *Ibid.*, p. 443.

Where practices of this sort involve combination with nonlabor groups the Hutcheson decision would seem to indicate that they are illegal under the antitrust laws. Where unions engage in such schemes alone there is still some question as to whether or not they would be exempt. This has led to some shifting of methods of procedure in order to accomplish the same purpose but maintain immunity under the law. Recent union action designed to bring about a coordinated increase in the price of bread is probably a case in point.¹⁹

The position taken by the members of the Antitrust Division of the Department of Justice seems to be essentially this: Labor does and should enjoy certain types of protection and immunity under the law so far as these are necessary to equalize the bargaining power of labor groups in "reasonable" actions intended to control wages, hours, and conditions of work and employment. Having been given this protection, through a long and gradual gain in the courts, and through the recent Norris-LaGuardia and National Labor Relations Acts, labor should also be held to account in practices that are clearly unreasonable restraints of trade. Labor should not be permitted to use coercive methods that place "unreasonable" burdens on employers and the general public, blocking progress and lowering planes of living. The attempt to define what should be regarded as "unreasonable" restraints has led to a specification of the five kinds of situations that have been listed here.

There are two major legislative changes supported by Mr. Arnold and included in some currently proposed labor legislation:

1. The first of these makes it a violation of the antitrust laws "to induce or require any employer to impose or adopt unreasonable restrictions or conditions upon the use of any material, machine, or equipment." This prohibition, however, is qualified by a proviso to the effect that restrictions or conditions of this nature shall not be deemed unreasonable "if they are directly and appropriately related to the wages, hours, health, safety, or working conditions of the employees, or if they are required in good faith as part of a temporary program to retard technological unemployment . . . or to forestall the imposition by the employer of a speed-up system."²⁰

2. The second is a series of provisions designed to prevent one union from attempting to break into the territory of a rival union. These provisions will be discussed in the following pages.

Evaluation of arguments of the opponents of the Antitrust Division's position in labor cases

There are a number of people and groups, especially the American Federation of Labor, opposing the recommendations for new legislation

¹⁹ An instance in the bakery industry was cited by Mr. Edwards. See the footnote on p. 482.

²⁰ Witte, Edwin E., "A Critique of Mr. Arnold's Proposed Anti-labor Amendments to the Antitrust Laws," *American Economic Review*, Vol. 32, No. 1, Part 2, Supplement, March 1932, p. 450.

that are supported by Mr. Arnold.²¹ We will consider their main points briefly:

1. The proposal that unions should not be permitted to "induce or require any employer to impose or adopt unreasonable restrictions or conditions upon the use of any material, machine, or equipment" is opposed primarily on two grounds: (a) It is argued that this would place too much power in the Antitrust Division in deciding what are "reasonable" practices, especially where problems of technological displacement of workers are involved. It is held that even though the final decision would rest with the courts, many cases would in fact be decided by consultation with the Antitrust Division and never be carried through to the courts, and that the power of that division would therefore be "too" great.²² (b) It is argued that the blocking of technological advance by the practices of business groups is quantitatively far more important than blocking by labor groups, and that these business practices would still go on unchecked by antitrust control.²³ This appears to be not so much an

²¹ Not all laborers or labor leaders are opposing the Antitrust Division's program. Members of the Antitrust Division cite literally thousands of letters from laborers commending their policies. The reason for this fact is well summarized in the following statement by Mr. Wendell Berge, Assistant to the Attorney General of the United States: "It seems to me that there are many other advantages to labor in our antitrust policy. The public will have greater confidence in labor unions if they are subjected to the same reasonable limitations as other groups, and if they eliminate those restrictive practices that have no reasonable relation to the promotion of better wages, hours, working conditions and the recognition of the principles and procedures of collective bargaining. Those who desire to prejudice labor in the minds of the public are constantly citing the occasional case of the employer prevented from using more efficient methods by the coercion of a labor group, the employer who is being held up by a jurisdictional strike, the union whose leaders have conspired with distributors to hold up prices, or the unions that have fallen into the hands of leaders who have conspired with employers for their personal gain or aggrandizement. I believe that the rank and file of labor are not interested in the perpetuation of these uneconomic practices. I believe that the cause of organized labor will be promoted if these practices are eliminated by the vigorous application of the antitrust laws." (Berge, Wendell: "Should the Antitrust Laws be Applied to Labor Unions?" An address presented at the Harvard Law School, March 28, 1940; pp. 28-29. Mimeographed.)

²² In an address to the American Labor Club on January 27, 1940, Mr. Arnold comments as follows: "The cry has been raised that the responsibility to determine the reasonableness of reckless conduct should not be on the Antitrust Division. This is exactly the same cry which accompanies enforcement against business or enforcement against agriculture. It is inherent in the antitrust laws; it is a necessary consequence of the application of the rule of reason that *someone* make a judgment on the legality of coercive activities of organized groups which restrain production and distribution. And furthermore, it is not the Antitrust Division which makes that judgment. The question is first submitted to a grand jury; it is then submitted to the United States District Court; next to the Circuit Court of Appeals; and finally to the Supreme Court of the United States. Nothing could be further from the dictatorial ruling of an individual than this process. It is the traditional judicial process under which our freedom is maintained."

²³ Mr. Witte expresses this opinion with some vigor: "The failure to utilize the cheapest materials, the best machines, and the most advanced techniques is much

argument against preventing labor restraints as an argument for revisions of public policy in business practices.

2. Provisions designed to prevent one union from breaking into the territory of a rival union took several forms. Each of these has been criticized on these principal grounds: (a) It is argued that attempts to control such actions under the antitrust laws would seriously handicap union organization in new areas since two new unions are very likely to spring up. This objection is countered by the arguments that the law would encourage labor to "get together with itself" to the benefit of all concerned. (b) It is argued that these provisions would block the development of new independent unions and thus remove a check on the action of leaders in the established unions. In reply to these two arguments the Antitrust Division states that the critics ignore the fact that the kind of policy recommended by Mr. Arnold would protect workers against unreasonable restrictions on their attempts to introduce new organizations, and that it is only unreasonable practices that would be illegal. Thus Mr. Arnold wrote:

From the workman's point of view the right of collective bargaining is itself in jeopardy if a powerful union may destroy a union which is the free choice of the workers in a particular plant.²⁴

And to this the labor leaders in turn reply that Mr. Arnold would in fact bring about conditions fostering and supporting the company union. (c) It is argued that the settlement of jurisdictional disputes should not be turned over to the Antitrust Division. But the Division would be concerned only with the enforcing of compliance with decisions reached in other ways, as has been stated explicitly by Mr. Edwards and others. The basic issue here goes back to the question of how these jurisdictional decisions are arrived at in the first place and whether it is practicable to treat them as contractual obligations and legally enforceable awards.

Conclusion .

Out of this mass of conflicting opinions a few general conclusions emerge:

1. There are many labor practices, such as strikes and boycotts, that interfere with interstate trade, but so long as these practices are conducted

more commonly due to the employers than to the labor unions. If working people are to be prosecuted because they place restrictions on the use of materials machines, and equipment which Mr. Arnold deems unreasonable, why should not the same treatment be accorded to employers who fall short of attaining maximum possible production because they, without reasons satisfactory to the Antitrust Division, fail to utilize the most advanced techniques or because they allow some or all of their equipment to lie idle? Certain it is that the loss to society by reason of the failure of business to operate always at full capacity has been many times as great as the one billion dollar cost per year which Mr. Arnold ascribes to the labor unions." *Op. cit.*, p. 453.

²⁴ Arnold, Thurman: Statement before the TNEC, *Hearings*, February 13, 1941. p. 6.

in pursuit of labor's "legitimate objectives" they would not be regarded as "unreasonable restraints" by either Mr. Arnold or his critics. Disputes may arise, however, as to whether particular acts under particular circumstances are in fact directed toward these legitimate objectives. For example, extremely high initiation fees, coupled with agreements for the closed shop, are tools for maintaining high wages by restricting the entry of workers into a particular occupation; are such policies "reasonable" or "unreasonable"? Mr. Arnold would probably say that they were not "reasonable" restraints involved in the pursuit of the legitimate end of raising wages, presumably because they did not directly involve wage adjustments.²⁵ A. F. of L. leaders would be very likely to disagree.

2. Opinions as to how far we should go in attacking labor restraints of trade under the antitrust laws differ; but all groups recognize that extreme cases of "unreasonable" restraints of the kinds listed on pages 500 and 501 are very "bad" and should be curbed in some way, though they may disagree as to the most appropriate methods.

3. There is still some doubt as to the legal position of trade-unions in cases of extreme practices that involve the fixing of prices to consumers but in which there is no proof of collusion with employers.

4. There is general agreement that practices in unreasonable restraint of competition in which there is collusion between laborers and employers are and should be illegal under the antitrust laws.

²⁵ See his comments on the \$500 initiation fee in his speech before the American Labor Club, *op. cit.*, p. 7.

CHAPTER 34

Legal Regulation of Wages and Hours of Work

LEGAL provisions for a minimum below which wages will not be permitted to fall, or for a maximum above which hours will not be permitted to rise, have results very similar to trade-union actions along these lines. But there are just enough significant differences so that we are not justified in merely dismissing these matters in terms of the generalizations reached in our discussion of trade-unions. We will therefore examine minimum wage legislation in some detail.

The History of Minimum Wage Legislation

By "minimum-wage regulation" is meant the fixing of a minimum level below which wages may not be permitted to fall. Such regulation may differ considerably in different times and places, according to coverage, the basis on which the minimum level is determined, and the purpose of the action. Coverage may be of particular industries only, of particular groups only (for example women), or it may be general. The rate set may be based on one of various kinds of estimates of need, or it may be that there will be an attempt to set it according to "value of service rendered." The purpose will usually be either to improve the living and hence the health and morale of the protected group, or to encourage recovery from a business depression by increasing the purchasing power of workers. What does the history of wage regulation in the United States reveal concerning each of these points?

Attempts of social reformers in the United States to get legislation with regard to wages, and of government to put such legislation into effect, have traveled a rocky road. While maximum hour legislation for women and minors and for workers in dangerous occupations was more readily accepted as a health measure, interference with wages was held to be dangerous tampering with the economic system, bound to precipitate serious unemployment among those for whose benefit the minima were set. Moreover, the emphasis on the rights of the separate states constituted a further barrier to the introduction of minimum wage legislation on a national scale.

State laws

The first efforts to set minima to wages were undertaken by the states, beginning with the Massachusetts law of 1912. This law provided for a

minimum wage commission that should establish wage boards in different industries as needed. Industries regulated were those in which there appeared to be "exploitation" of low-paid labor of women and minors. The boards were to set rates on the basis of need to provide "a living wage." The general plan followed the precedent set by the British law of 1909; but enforcement powers were publicity only. By 1923 fourteen states (mostly agricultural), Puerto Rico, and the District of Columbia had followed the lead of Massachusetts. In some of these there were more powers of enforcement, but only the relatively weak laws survived the famous Adkins case decision in 1923.

In the Adkins case the Supreme Court invalidated the District of Columbia legislation as unconstitutional. It had been defended in the court as protecting the health, safety, and morals of the people. Justice Sutherland wrote the opinion declaring these laws unconstitutional, as interfering with "property" rights of employers. The argument of the Court was based in part on the fallacious assumption that individual wages are without exception equal to the "value of service rendered," that is, the marginal revenues obtained from the workers employed.

The effect of Justice Sutherland's decision was to cause a shift in the formulation of subsequent laws. A way out of the constitutional difficulty was seen in basing the determination of the minimum level in part on "value of service" instead of merely on need, providing that rates set should be "commensurate with value of service or class of service rendered." Wages in particular industries thus fell under the law not merely because they were low, but because there was also monopsonistic control that succeeded in depressing wages below marginal revenues. The approach was still industry by industry, and the purpose was still the providing of better living for the protected workers. Wisconsin framed a law of this sort in 1924, but little action was taken elsewhere until the depression of the 'thirties. The year 1933 brought seven new laws, all of which included a "value of service" clause, and in 1934 a mandatory clause was added to the old Massachusetts law. There are now at least 25 states having such laws. Oklahoma is the only state to have brought men as well as women within the scope of this kind of law.

Interstate pacts

A state attempting to set minimum wages above those paid in surrounding states is at a disadvantage from several points of view. Higher wages in the regulated industries will push up costs, and if these industries were just covering costs, they will now find themselves unable to compete with rivals in other states where wages have not been regulated. The reduction of employment that would follow in the state having the wage regulations may therefore be important, whereas it might have been insignificant had all states come under the same regulations.

A recognition of the difficulties attendant upon setting minimum wages in one state but not in neighboring states led, in 1931, to an interstate conference in the east instigated by Franklin Roosevelt—then governor

of New York. Conferences were held until in May 1934 seven states (Connecticut, Maine, Massachusetts, New Hampshire, New York, Pennsylvania, Rhode Island) signed a pact aimed to secure uniformity of standards for conditions of employment, especially wages and hours for women. The New York law was taken as the standard and by 1938 Massachusetts, New Hampshire, and Rhode Island had ratified the agreement and Congress had given its consent.¹

Federal action

There are two ways in which greater uniformity of wage standards may be sought. One is the extension of interstate pacts in different regions, the other direct federal legislation. Those concerned in maintaining a maximum degree of local self-government, the defenders of "states rights," prefer to emphasize the former. The interstate compact has the advantage that states in each region can legislate for conditions in that particular region. There is no problem (as in the case of federal legislation) of deciding whether to set uniform money wages for all regions or to adjust minima by regions to differing costs and patterns of living. Interstate pacts have the further (and as we shall see very important) advantage that coverage within states can be much more nearly complete and the interstate pact can focus on conditions in particular industries, whereas it is constitutionally impossible for federal legislation to accomplish this. The federal government can set wages only for those workers who are included under the interstate commerce clause of the Constitution and for workers employed by the federal government or employed in the production of goods purchased by the federal government. A federal law on the other hand can be accomplished more quickly than can interstate agreements. If we want the kind of control the federal government has power to impose, then this method should be encouraged. Whether or to what extent we want these controls is another question.

Attempts of the federal government to control wages have been of three main types. The first was indirect control through cooperation with the representatives of each industry under the National Industrial Recovery Act in 1934. (Prior to this there had also been an attempt to control wages indirectly through controlling hours of work and "overtime" pay on the railroads in 1916.) Although some of the wage agreements arrived at under the NRA codes had the effect of raising wages of low-paid labor in monopsony situations, they served primarily as a support to demands of trade-unions to raise wages generally whatever the level from which they started. The bases of wage setting were neither need nor value of service, although they were more nearly compatible with the value of service idea. The purpose was to raise wages generally in order to expand purchasing power and hence increase business activity. That this reasoning was in large measure fallacious will be argued in a subsequent section of this chapter.

¹ Vyver, Frank T., "Regulation of Wages and Hours Prior to 1938," in *Law and Contemporary Problems*, Vol. VI, No. 3, 1939, pp. 326, 327-331.

Second, attempts at wage setting from 1934–38 by the federal government in its position as buyer and employer concentrated on government adherence to “prevailing wages” and had the effect of pegging these wages. Thus the United States Housing Act requires that “prevailing wages” be paid on jobs done under that act. These “prevailing wages” are usually union wages in union-restricted occupations, and in effect the Act prevents competitive wage cutting. The Walsh-Healey Act of 1936 requires that “prevailing wages” be paid on work the products of which are purchased by government. Administrative decisions as to what “prevailing wages” are usually accept union scales where these exist. Insofar as government buys from industries characterized by “sweated” conditions it is likely to set wages higher than those actually paid in these industries and may, by providing opportunities for workers at better pay, have the effect of pushing these wages up.

Third, the federal Fair Labor Standard Act popularly known as the federal “Wages and Hours Law” was passed in June 1938. This was the culmination of a long and persistent drive to regulate minimum wages and maximum hours, which had taken on an intensity following the termination of the NRA. The provisions of the Act are as follows: ²

It sets up a rigid scale of wages and hours for workers in interstate commerce or who produce goods for such commerce. A statutory minimum of 25 cents per hour is in effect the first year, then a 30-cent minimum is the standard until October, 1945, when the 40-cent wage becomes operative (Section 6). Unless there is payment of one and one-half times the regular rate of pay covered employees may not be worked in excess of 44 hours per week the first year, 42 hours the second and 40 after the second (Section 7). The enforcement of the Act is under the control of a single Administrator in the Department of Labor. Industry committees, appointed by him, may make recommendations that can be made the basis of a wage order setting a minimum wage in excess of the universal rate but not more than 40 cents an hour (Section 8). There are numerous exemptions, some outright, some dependent upon administrative regulation. There may be criminal and injunction proceedings against violators and civil recovery of double damages by workers (Section 16). Child labor under 16 years of age is prohibited and may even be eliminated in certain occupations up to 18 years of age (Sections 3 (1) and 12).

The coverage of the Act is as general as was administratively practicable and constitutionally permissible; particular industries are omitted as exceptions, instead of being included as especially needing regulation. Since this is a federal law, its coverage is limited to workers engaged in interstate commerce or performing activities necessary to such subsequent commerce. The lowest wages and worst working conditions are probably most likely to appear in occupations in intrastate commerce, so that the law misses some of those situations that are generally regarded as most needing regulation. The basis of the determination of what should be the minimum level of wage permissible is essentially estimates of need. The

² Forsythe, John S., “Legislative History of the Fair Labor Standards Act,” *Law and Contemporary Problems—The Wage and Hour Law*, Vol. VI, No. 3, Summer 1939, p. 473.

purposes of the act were mixed. It was urged both by those seeking primarily to raise the level of living of workers in the lowest-paid occupations and by those who argued that higher wages would help bring back prosperity.

Analysis and Evaluation of the Effects of Minimum Wage Regulation

In analyzing and evaluating the effects of minimum wage regulation, we shall confine ourselves to five main questions: (1) What are the effects of setting minimum wages above previously existing levels where labor markets are purely competitive? (2) What are the effects of setting minimum wages above previously existing levels where there is monopsony in the buying of labor? (3) What are the effects of minimum wage regulation on management efficiency? (4) Is there any validity to the oft-quoted argument that "a minimum wage tends to become a maximum"? (5) What validity is there, if any, in the "purchasing power" argument for regulating wages in order to bring prosperity in general business?

Some effects of minimum wage regulations under purely competitive conditions

The core of opposition to minimum wage regulation has been the argument that any interference with the operation of the competitive mechanism was a dangerous and undesirable procedure. It is argued that that mechanism, when pure competition is universal and there are no restrictions on entry, operates in such a way as to allocate resources where they are most wanted, to lead workers into the jobs in which they would receive the highest pay, and to adjust that pay so that in each case it equals the full value of the service added. Intervention to raise wages would in this case cause unemployment, which was considered a sufficient reason for opposing government intervention. Furthermore, it was assumed that the economy was essentially purely competitive throughout and that restrictions on entry were the unusual exception. The effects of wage legislation in a purely competitive economy would be, briefly, as follows:

1. *We cannot avoid the fact that even in a purely competitive economy there is no guarantee that everyone can find work in which he can contribute enough to the economic product to earn a decent living wage.* Some workers have obligations greater than others. Some are of such poor calibre that they are not capable of contributing in the productive system enough to maintain a family or, in some cases, even themselves alone; society takes some responsibility in the more extreme cases. Some are supported entirely by relief; others receive relief subsidies as supplement to small pay. *A minimum wage higher than what these workers can produce would throw them out of work.* They would then be completely dependent on government relief, which might or might not involve work for the government. Evidently society will not in any case permit compe-

tition to leave its submarginal victims along the way. The question resolves into one of *how* government should act, and whether submarginal workers will be unemployed, employed by government, or employed by private enterprise with government relief subsidy.

2. *Unfortunately the present federal law may even exaggerate the difficulties of the lowest-paid groups, because they are often in uncovered occupations and these occupations are all that remain open to some of the workers forced out of employment where wages are regulated.* The increased numbers of workers then seeking work in these uncovered areas forces the wages there down to levels still further below the standard provided in the law.

3. *Minimum wage legislation may in some cases force more rapid shifting of population from one area or occupation in which wages are exceptionally low to another in which the opportunities are greater. It does this by forcing unemployment in the poor area or occupation.* A more direct and less costly way of attacking this problem would be the expanding of services of employment agencies, both public and private, and the elimination of the "bargain basement" employment agency trading on distress cases. Such an expansion would assist both workers and employers to make informed decisions.

4. To some extent wages in underprivileged areas may be increased by an inward movement of industry when labor fails to move out. This change in the relative supplies of material agents and labor available will raise the marginal revenues obtainable from any given group of workers in the underprivileged area. Thus the movement of textile mills from New England into the Southeast has tended to raise wages of workers in the South (provided labor is hired competitively). *A premature legislative raising of the minimum wage for Southern workers might check the movement of capital into the area and thus prevent productivity from increasing.*

5. Even in a purely competitive regime, incomes are unequal to a high degree. *Would the general establishment of minimum wages be a feasible means of modifying this inequality by jacking up the lowest income level?* It may do so to a very limited extent by (a) shifting incomes from higher-paid workers to those who were affected by the wage regulations and are now either unemployed on public relief, or employed at the new higher rates, or (b) shifting incomes from property-holders to the protected groups of laborers. On the other hand the unemployment entailed in this program may bring exaggerated problems of low incomes and hence of inequality between low and high income groups.

6. Minimum wage legislation affects consumers primarily through its effects on the total output of the economy.⁸ Insofar as it causes unemployment, resources are idle; the national real income is therefore lowered. On the other hand, guaranteeing a certain minimum wage may prevent the deterioration of individual workers otherwise on very low levels of

⁸ Effects on resource allocation are probably of minor importance.

living; over a long period of time the result may tend to be a more productive labor population and hence a higher national plane of living.

Minimum wage regulation in an employer-monopsonized labor market

In the discussion in the previous section we assumed the labor market in each industry and locality to be predominantly competitive. But a variety of conditions may, taken in conjunction with immobility of labor out of certain industries or areas, lead to strong employer-monopsony or oligopsony in some labor markets. Those arguing for minimum wage legislation have often emphasized the fact that not only are some sectors of the labor market controlled by one or a few employers, but also that laborers in these situations are often unable to organize themselves to form a counter oligopoly in the sale of labor. When an employer buys monopsonistically labor that is offered competitively he will be able to push wages down by restricting his inputs somewhat, and it will generally increase his profits to do so. A minimum wage higher than what he had been paying will remove the lower-wage, smaller-working-force, alternative. Since he must now pay the legal rate anyway, he will find it advantageous to hire more men. He will add men so long as the legal rate is less than the marginal revenue of the workers employed. In fact he will aim to add men until wages and marginal revenue are equal (provided enough workers are available at the legal minimum). This is just the same situation as that we described in discussing the effects of trade-union wage regulation in the face of monopsony.

Minimum wage regulations and management efficiency

Thus far we have assumed that the employer who finds himself in a position to depress wages will inevitably do so in a degree consistent with his all-encompassing goal of maximizing profits. Actually many employers do not squeeze workers even when it would be to their financial advantage to do so, but others push down wages and fail to develop other economies. An advantage of minimum wage regulation that has been argued strongly by some of the more progressive industrialists is the effect that enforcing higher wages may have on efficiency of management in industry. Mr. Filene urged this point emphatically and drew on the conclusions of Hoover's investigation of Waste in Industry to prove his point. Inertia and the attempt to get along by paying low wages has characterized many entrepreneurs who could easily save much by more efficient management. Such producers have of course not been maximizing their profits in the past, but the greater urgency of the situation when labor costs rise may force them into finding neglected wastes and eliminating them. To the extent that this occurs—and there is no doubt that it is important—no unemployment may be caused by the increase in wage rates, and the income of the society as a whole may be increased. Labor will receive a larger share of a larger total.

Consideration of the argument that "a minimum wage tends to become a maximum"

Trade-unions have sometimes opposed minimum wage regulation, on the grounds that "a minimum wage tends to become a maximum wage." This is an extreme statement. What is usually meant is that the setting of a minimum wage may lead to an adjustment of the wages of those receiving higher pay down toward the minimum. The degree of truth in this statement will depend on particular circumstances and on the level at which rates are set. We shall confine our analysis to a national legal minimum.

First, if minimum wages are introduced in a perfectly competitive situation they can have no effect unless they are set higher than competitive rates. Wages would not be pushed below the competitive level so long as entrepreneurs continued to compete for workers.

Suppose a national legal minimum were set higher than competitive equilibrium rates for unskilled labor. What would happen to skilled labor would depend on whether it was substitutive for or complementary to unskilled labor. If it were substitutive, the opportunities open to skilled workers would be improved and their wages would rise. If, on the other hand, it were complementary to unskilled labor, skilled labor would find its employment opportunities less favorable than previously, and wages for skilled labor would then fall. Whether skilled labor was substitutive for or complementary to unskilled would depend on the particular technological possibilities and the prices of cooperating material agents in each industry.

The argument of the trade-union leaders that "a minimum wage tends to become a maximum" is usually made on the basis of various assumptions about the noncompetitive aspects of wage determination. Just as entrepreneurs may ride along neglecting wastes of inefficient management until jolted by an increase in labor cost, so they may also have neglected the possibility of reducing the wages of skilled labor. A minimum wage law may push them into squeezing their costs in this way. But of greatest concern to union leaders are the effects of minimum wage legislation on the relative bargaining positions of workers and employers. Such legislation may affect labor bargaining in either direction. It may strengthen the position of the weaker bargainers and lower-paid workers, giving public support to their demands in both regulated and unregulated industries. It will also strengthen their position in bargaining on other issues by relieving the cost of pressure on the wage question. But strong unions bargaining for higher wages are likely to find themselves in a less satisfactory position. The difference between their wages and the minimum rate will be sufficiently clear so that public support of their demands is less enthusiastic. Employers of such labor may in the same way find their positions strengthened. The fact that the A. F. of L. has strongly opposed wage legislation for men while the C.I.O. has favored such legis-

lation is a significant commentary on the differences in membership of the two organizations.

Evaluation of the argument that higher wages may help bring prosperity

One of the arguments most persistently urged in recent years by those favoring federal minimum wage regulation is based on the proposition that higher wages would mean greater income, increased spending, and hence a return to prosperity. A full consideration of this question would carry us deep into the disputes concerning business cycles, and cannot be undertaken here. A few brief comments will, however, suggest some of the elements in this problem. First, it is obvious that an increase in wage rates could increase labor income only if the demand for labor is inelastic; otherwise there may be extensive substituting of material agents for labor, or an immediate reduction in production. Higher wages mean higher costs and production will be stimulated by these wages only if prices rise faster than costs. If the higher wages give more income to labor as a whole, the savings of capitalists may be diminished, the total amount spent on consumers goods increased, and production of these thus stimulated. Even if such a result were to ensue, the long-run effects of a persistent policy of minimum wage regulation have to be considered. The argument that setting minimum wages will help bring prosperity stands on very shaky grounds and is a dangerous one. It would tend to lead to higher minima than would the other motives of regulation. The ramifications of this argument will be considered further when we study business cycles.

Some Practical Difficulties in Setting Minimum Wages

When the various theoretical possibilities that we have discussed are surveyed, it must become evident that the problem of establishing an "ideal" plan would be an exceedingly difficult one. To whom will these minimum rates apply, under what circumstances will they be invoked, what level or levels of rates will be set, how will other government policies such as the giving of relief be coordinated with the wage controls? These and many other questions arise. An attempt to discuss with any degree of completeness the practical difficulties involved in these programs would take us into many more pages. We shall confine our comments to two main points: (1) the problems involved in establishing rates consistent with a "minimum of health and decency for all"; (2) some of the difficulties involved in the problem of coverage.

Determination of the base rate

It is generally assumed that the American economy is capable of providing for the entire population means of living that would meet acceptable minimum standards. It is very difficult, however, to decide what these minimum standards should be. Costs of living in different areas vary. Should these differences be taken into account in setting

rates? Should long-established customs be considered evidence of the essentiality of a good or should minimum standards be set in which the wage is just sufficient to cover some scientifically planned minimum budget? Should the low standard of some groups such as those of back-eddies of Southern whites in the mountains of Kentucky be accepted for these workers when the workers themselves clearly show their disinterest by working less when pay is raised? If low rates are set for these workers should these rates be universalized as legal minima?

In coordinating social policy the question will inevitably rise, should the minimum standard for those on relief be the basis for the minimum wage rate or should the latter be higher? If the two are different there is room for some question as to what society intends, is it setting a higher than minimum standard for wages? a lower than minimum for relief clients? On the other hand, relief payments as high as wages may discourage active search for work. Relief standards are relatively elastic, based on family need, and do not have to meet the problem of consistency in individual cases with value of service rendered.

The needs of large families raise further problems. While the economy may be capable of providing for all its members a decent plane of living, it does not follow that it could provide for every worker enough to support a family of five. Yet many laborers do have to support such families on their earnings. If wages were set with minima based on individual need, the most needy would remain unemployed since if it were necessary to pay a higher wage to the family man he would be left out. Yet to set general rates covering adequately his needs might be unsupportable and lead to extensive unemployment. This dilemma has been solved in some countries by resort to a system of family allowances. A minimum is set for all workers on the basis of the needs of the single man or woman, this income is then supplemented by family allowances from some centrally established fund. The fund may be in the hands of the government, derived from the community by taxes of one form or another. It may, as has occurred in some European countries, be established by industries. In these cases all employers contribute a certain amount per worker to a central industrial "equalization fund." From this fund supplementary allowances are paid to married workers with dependents. The system is not forced to provide to all a family wage and hence support a lot of hypothetical wives and children. If a state equalization fund is obtained by taxing high incomes it becomes a means of equalizing incomes. If it is derived, as in the European industry schemes, from what amounts to a tax on pay rolls, then the family allowances become chiefly a means of redistributing incomes among workers.

When a reasonable minimum rate has been decided upon, when group and area differences have been considered, and when the matter of family allowances has been given attention the chief problems of determining "need" have been taken into account. The minimum rate is not yet determined, however, until its feasibility in relation to the economic contributions of the various workers has been considered. To what extent should

considerations of the low marginal revenues obtainable from some groups be taken into account to modify the rate set, or as a basis of exceptions? Will high rates bring extensive unemployment?—evasion with special pressure on the weakest individuals? These questions can be answered only jointly with consideration of government relief policies and government provision of employment for submarginal workers.

The problem of coverage

Many other issues that cause difficulty are arising with regard to coverage. These include questions as to where enforcement would be practicable. It would be extremely difficult to control wages where board and room are part of the pay and where there are many small employers, as in domestic service. Unfortunately many of the "sweated" industries giving out home work and paying very low wages are difficult to regulate.

The American form of government with its emphasis on states rights has discouraged uniform national policy. States pursuing different policies run into problems of interstate competition, and the tendency for labor to move in and business to move out of a state when minimum wages are set there but not elsewhere. Since federal regulation is tied to the interstate commerce power, employees in intrastate businesses or occupations will be excluded from federal jurisdiction. Minimum wages for these groups are set only by state action.

Especial attention should be given to the fact that incomplete coverage may actually aggravate the conditions which the wage legislation is intended to amend. This is likely to be true in part of the recent federal wage and hour law. That law hardly touches workers in the sweated industries. Insofar as it raises wage rates elsewhere it is likely to cause some discharge of workers, who then swell the ranks of those working for yet lower pay in industries immune from regulation.

History of Legislation Controlling Hours of Work

Ever since the end of the eighteenth century there has been agitation for the shortening of hours of work, and this became increasingly important as machine industry and routinized production occupied a greater and greater place in the total of economic activity. Efforts have been made to get shorter hours both through trade-union bargaining and through legislation, especially since 1860. The ten-hour day had been very generally adopted, voluntarily or otherwise, by 1890. There are some examples of persisting long hours of work, however, that would surprise many people today. For example, cotton manufacturing, iron and steel plants, and bakeries continued a working day ranging from eleven to thirteen or more hours with night work, and in some cases there was also a seven-day work week. It was as late as 1923 that the steel industry gave up a twelve-hour day, seven-day week for the majority of its workers.

Two major purposes have lain back of agitation to shorten hours through social legislation. The first of these, predominant in such agitation until very recent years, was the importance of shortening hours in

order to improve the health, morale, and efficiency of the workers. More recently emphasis has been to a much greater extent on the use of maximum hour legislation as a means of forcing a sharing of job opportunities in the face of serious unemployment.

Until 1938, legislation setting maximum hours was almost entirely state legislation. Exceptions were the railroads, and the hour provisions in the codes set up through joint cooperation of representatives of industry and government under the NRA. Such legislation applied to special classes of workers only. At first women and children only were included, though in part with the intention of establishing regulations for women that would lead to a voluntary generalization of these maxima to apply to men also. New England led these developments, with legislation for women and children in textile factories in the middle of the nineteenth century. Next followed legislation for workers in especially hazardous occupations such as mining. More recently two other groups have been included in both state and federal legislation; government employees and those engaged on government work, and railroad workers and others in occupations affecting directly public safety. Finally, the "Fair Labor Standards Act" of 1938, described earlier in this chapter, covers most workers engaged in occupations that may be regarded as closely enough related to interstate commerce to come under federal control.

Part VII

THE DISTRIBUTION OF INCOME

PREAMBLE

MR. YANG was from China, and he wanted to know what America was like, so he was put into the hands of an able guide who knew the country well.

The guide took him first to see the Rockastorbilts. They live in a big house in the city and several big houses in the country. They drink fine old imported wines, and eat fresh broiled lobster. They wear rubies and diamonds and fine Scottish tweeds. Mr. Rockastorbilt is driven to the office in a big black car by a swank chauffeur, and Mrs. Rockastorbilt is driven to Bergdorf Goodman's on Fifth Avenue in the same way, while young Mr. Rockastorbilt takes out his "Dates" in a trim canary yellow "sport job" running on twenty cylinders. The Rockastorbilts collect paintings, and attend the opera. They vacation at Palm Beach or on the French Riviera.

Mr. Yang was amazed. He asked: "Is this the way Americans live?" And the guide replied: "No, this is the way a very few Americans live." So they went on their way.

The guide next took Mr. Yang to see the Patronicavettes, who live in the slums of a big city—and to see the Williams family, who live a hundred miles from nowhere, on a little farm that produces hardly anything at all.

There are ten people in the Patronicavette family. They live in a tenement that has one window at the front and one at the back, with no windows at all in the two rooms between the front and the back. They share a hall watercloset with another family like themselves. They eat bread, and potatoes, and soup with a little meat flavoring, and dried beans—though they do not always eat. Their clothes are patched, and some of them do not have any shoes at all. The children play in the alley, Mr. Patronicavette chums around with cronies at the corner, and Mrs. Patronicavette gossips with neighbors on the broken wooden steps.

There are twelve people in the Williams family; they live in a two-room cabin with three beds, and lug their water to drink and cook and bathe. Mrs. Williams cooks in a big kettle, making soup and corn meal from the corn in the little patch by the house, adding a little salt pork flavoring and

some dried beans. When the children are hungry they suck their fingers and their legs grow crooked.

Mr. Yang was disturbed: "Is this, then, the way most Americans live?" The guide replied: "No, there are more Americans who live this way than who live as the Rockastorbilts; but most Americans do not live this way." So they went along.

They came to the Smiths, who have two children. The Smiths live in a small house almost (but not quite) like a lot of other small houses on the same street. The house is a little out of repair and does not heat very well, but it has running water. They eat pork chops, and frankfurters, and pot roast, and potatoes. They do not buy much in clothes; but the women manage to keep in style. Part of the time they have a car. The whole family goes to the movies. The guide said: "This is the way most Americans live."

Mr. Yang was puzzled. He wanted to know why there were families living like the Rockastorbilts and the Patronicavettes and the Smiths. So the guide explained:

He said that Mr. Rockastorbilt was clever at organizing big business enterprises—that the Rockastorbilts owned a lot of property in investments in mines and factories and shipping companies—and that they had a lot of stock in the Aluminum Company of America, which was a monopoly. He said that Mr. Patronicavette had been to school for only three years and did not have any job most of the time, though he sometimes got work loading and unloading ships, or cleaning the city streets. He said that Mr. and Mrs. Williams just got along on what they could grow on the eroded soil of the little farm they rented at \$10 a year from the lumber company that had cleared out the last timber five years before. He said that Mr. Smith did not have any property either, that he worked in a shoe factory. He said that some of Mr. Smith's neighbors were salesclerks and garment makers and plasterers and steel workers, and some of them ran small shops of their own. He said that every once in a while there was a lot of unemployment.

But Mr. Yang was not satisfied. He wanted to know whether people in America thought that this was the way things should be, and whether the government was doing anything to change the situation. So the guide replied again:

He said that there was a lot of disagreement about how incomes should be divided and what should be done about changing any distribution of income that seemed to be bad. He said that most people thought a man ought to get an income according to what he contributed to production, but they disagreed about just what they meant by this. He said that most people thought something ought to be done about the Patronicavette family and the Williams family, and that most people thought the government ought to collect heavy taxes from the Rockastorbilts. And he said that the government did do something about these things, though many people thought there was a lot more to be done.

At this Mr. Yang was both puzzled and unsatisfied, so he decided to examine the problem further.

Income distribution by family units

To the person receiving a yearly income of \$200 (say as a member of a family of five in which family income is \$1,000) it is small consolation that the average per capita income in the United States in 1940 was about \$550, or that the United States ranks second to the top in average level of living. The welfare of the people of a country depends fully as much on how the national income is divided as on the average per capita income. Are we a nation of many very low income individuals and families and a few people living in the height of luxury, or are we a nation in which most of the people share in fairly equal proportions the society's goods?

Typically, incomes are received and spent by family groups; and since most people live in family groups, income studies have frequently focused on comparisons of family incomes rather than per capita incomes. About one in every thirteen people receives and spends his income as a single individual; such people will be regarded here as "one-person families." The family, or income-receiving and spending unit, may then be termed a "consumer unit." Some of the most important facts about income distribution are facts concerning the proportions in which different consumer units share in the total national income. This is the first question we shall pose in examining the distribution of consumer incomes.

In the year 1935-36, national money income was roughly \$60,000,000,000. Population was about 128,000,000, divided into about 40,000,000 consumer units, or an average of slightly over three persons per consumer unit.⁶ On this basis, income per consumer unit was about \$1,500, but this average reflected heavily a few very large incomes, just as did the per capita figures already cited. This is indicated by the fact that an income of \$1,070 (the median) divides the total number of consumer units in half. Only 50 per cent got as much as \$1,070; the other 50 per cent got less. The figures are even more striking if income receivers are divided into thirds. Roughly one third of all consumer units received incomes of less than \$780. The middle third received between \$780 and \$1,450. And only the upper third received over \$1,450. It should be remembered that the average consumer unit comprises over three persons.

In the year 1941 money incomes were higher than in 1935-36, though prices had of course risen also. An income of \$1,481 (the median) divides the 1941 consumer units in half. The lowest third includes those with incomes up to about \$950; the middle third range from \$950 to \$2,100. In 1941 the upper third received over \$2,100.

⁶ All estimates given here exclude persons living in institutions such as military camps, poor farms, prisons, etc., totaling about 2,000,000 persons. Exclusion of this group does not greatly affect the estimates given. The source of 1935-36 data is the National Resources Committee, *Consumer Incomes in the United States*, U. S. Government Printing Office, Washington, D. C., 1938.

528 THE DISTRIBUTION OF PERSONAL INCOMES

Figure 35—1 shows the estimated distribution of income in 1935–36 by consumer units. The bars at the left indicate the per cent of consumer units in each income range, those at the right show the per cent of aggregate income received by consumer units in each income classification. For example, the greatest number of consumer units received \$500–\$1,000, though this group obtained only about 13 per cent of all income.

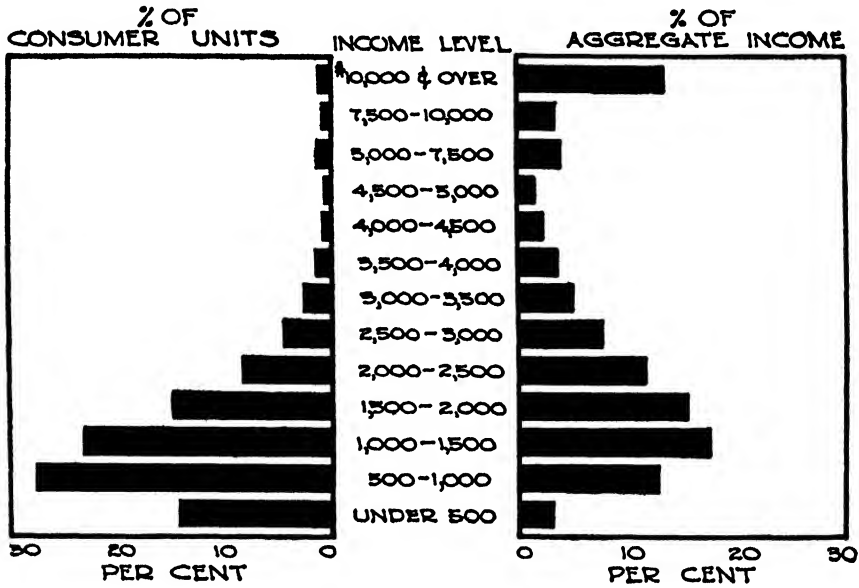


Fig. 35—1. Distribution of family income in the United States by income levels, 1935–1936.*

* Based on data from *Consumer Incomes in the United States*, Table 2, p. 6.

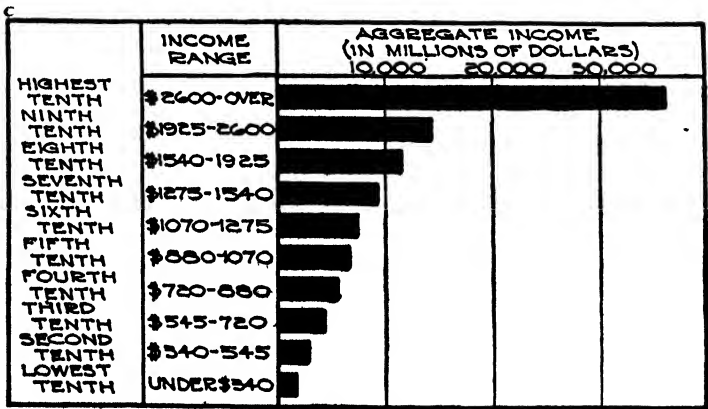


Fig. 35—2. Share of aggregate income received by each tenth of the nation's consumer units, 1935–1936.*

* Based on data from *Consumer Incomes in the United States*, Table 1B, p. 96.

Figure 35—2 arranges these data in another way, giving the aggregate total income received by each tenth of consumer units. It shows that only one tenth of all consumer units received as much as \$2,600, but this upper tenth received an aggregate income of \$21,452,000,000, or considerably over one-third the total national income. By contrast, the lowest tenth, getting under \$340 per consumer unit, received only about \$1,000,000,000, or around 2 per cent of the total national income.

This picture would of course be slightly modified for 1941. Figures are not available for as detailed an income classification, but a broad comparison of the per cents of consumer units in different income groups is shown in Table 35—2.

TABLE 35—2

DISTRIBUTION OF INCOMES IN THE UNITED STATES
IN 1935-36 AND IN 1941

Income level	Per cent of consumer units	
	1935-36 *	1941 **
Under \$500	17	16
\$500-\$1,000	30	19
\$1,000-\$1,500	22	16
\$1,500-\$2,000	13	14
\$2,000-\$3,000	11	20
\$3,000-\$5,000	5	10
Over \$5,000	2	5
	100	100

* Data are from *Consumer Incomes in the United States*, p. 6.

** Data are preliminary estimates by the U. S. Bureau of Home Economics and the U. S. Bureau of Labor Statistics, based on the "Small Sample" taken in 1941.

Income distribution by occupation

The occupation we follow is a basic factor determining the incomes most of us receive, though it is by no means the only factor. Data on incomes by occupational groups in 1935-36 are given in Figure 35—3, which shows the percentage of families in each occupation by income levels. For example, about 28 per cent of wage earners got less than \$1,000, about 35 per cent between \$1,000 and \$2,000, and so on. These figures show that on the average the lowest general income groups were farmers, wage earners, and independent businessmen (proprietors of relatively small enterprises, such as retail stores or very small factories). Those receiving the highest incomes were independent professional people (doctors, lawyers, and so on) and salaried businessmen. These figures, however, are only rough indicators. Especially in the case of farmers, noncash incomes (products used directly from the farm) may be an important supplement to the money incomes indicated here. Moreover, the incomes received by farmers rose much more sharply than most incomes between 1936 and 1941.

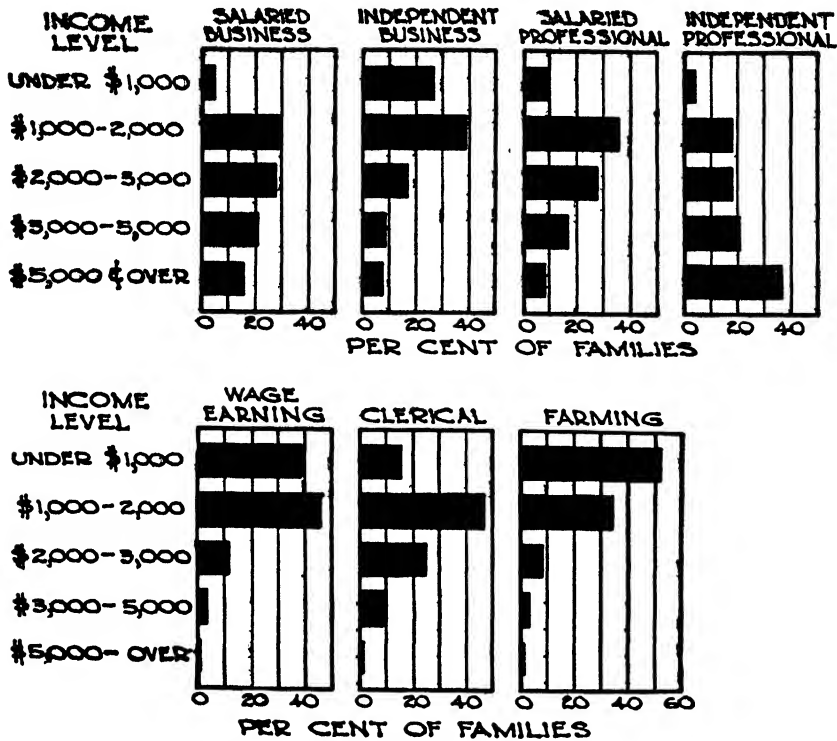


Fig. 35—3. Income distribution of nonrelief families in seven occupational groups.*

* Based on data from *Consumer Incomes in the United States*, Table 10, p. 26

Income distribution by area

Over the last decade the opening of the "Old South" and the remote mountain areas of the Cumberland and Piedmont plateaus in Kentucky, Tennessee, Virginia, and North Carolina has brought to the attention of increasing numbers of people in this country the extremely low level of living of considerable groups of the population. This increased awareness has in part accompanied the increased intervention of the federal government in the problems of agriculture; and it has found expression in the development of a Farm Security Administration to aid low income farmers and in the partial reconstruction of underprivileged areas under various types of public works programs. The problem of low incomes in the Southeastern mountains has been further dramatized by such literary efforts as "Tobacco Road," with its record run in the New York theatres.⁷

Many of the same developments have led to an increased appreciation of the low levels of living of many families in the Southwest; and simul-

⁷ In such dramatization there is also much danger because of misleading exaggerations of social degeneration which lead to increased prejudice.

taneous drought in the dust bowl and mass migrations have been so serious as to force an awareness onto many people who would have preferred to let "well enough alone." Here again, literature has contributed to popular comprehension. Steinbeck with his "Grapes of Wrath," quickly interpreted in the movies, has reached thousands of comfortable people who would otherwise have heard little of the tragedy of those other thousands of destitute migrants.

The map presented in Figure 35—4 shows per capita incomes by area in 1929. The low income level of the Southeast stands out sharply; so also do the relatively high average incomes in the major industrial states

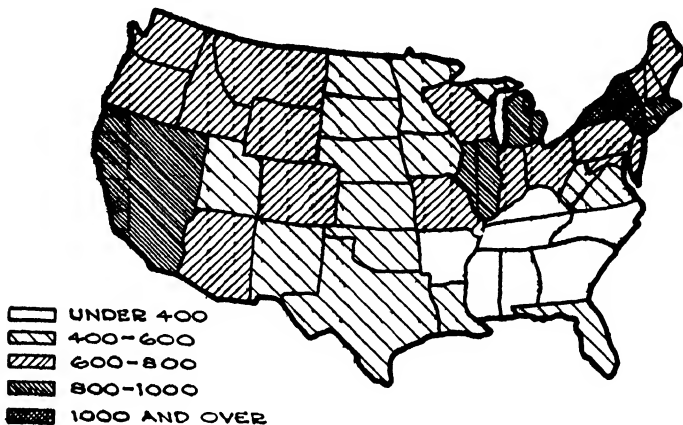


Fig. 35—4. Per capita income by states in 1929.*

* Based on Maurice Leven's estimates in *America's Capacity to Consume*, p. 173. (The figures utilized include profits from the sale of property and imputed rent on owned homes.) Published by the Brookings Institution, Washington, D.C., 1934.

of the Northeast, in some of the relatively sparsely inhabited ranch and mining country of the Rocky Mountains, and the rich lumber country of the Northwest. The critical conditions in drought-stricken plains states from North Dakota south to Oklahoma are of more recent development, though even in 1929 these were relatively low income areas. The extent of the differences in average incomes in different parts of the country will appear astounding to those who have sat back comfortably satisfied with the generally "high American plane of living." Even a quick glance at a map like this shows clearly that there is no such thing as "the American plane of living"; that there are many different planes of living among major groups of American people.

Age, sex, and race differences in income levels

There are many other ways in which people might be classified in a study of income distribution.

Young people receive much lower incomes than those in middle-age groups, but the peak income age varies considerably depending on occupation. In most occupations, the income peak comes somewhere

in the 'forties, but for highly skilled workers, business and professional men, the peak often comes later. Women receive lower incomes than men, averaging less than 60 per cent as high. This differential comes largely as a result of the occupations frequented by women and the fact that most women are employed only in their youth, usually leaving work before they have attained their full earning power. Women do

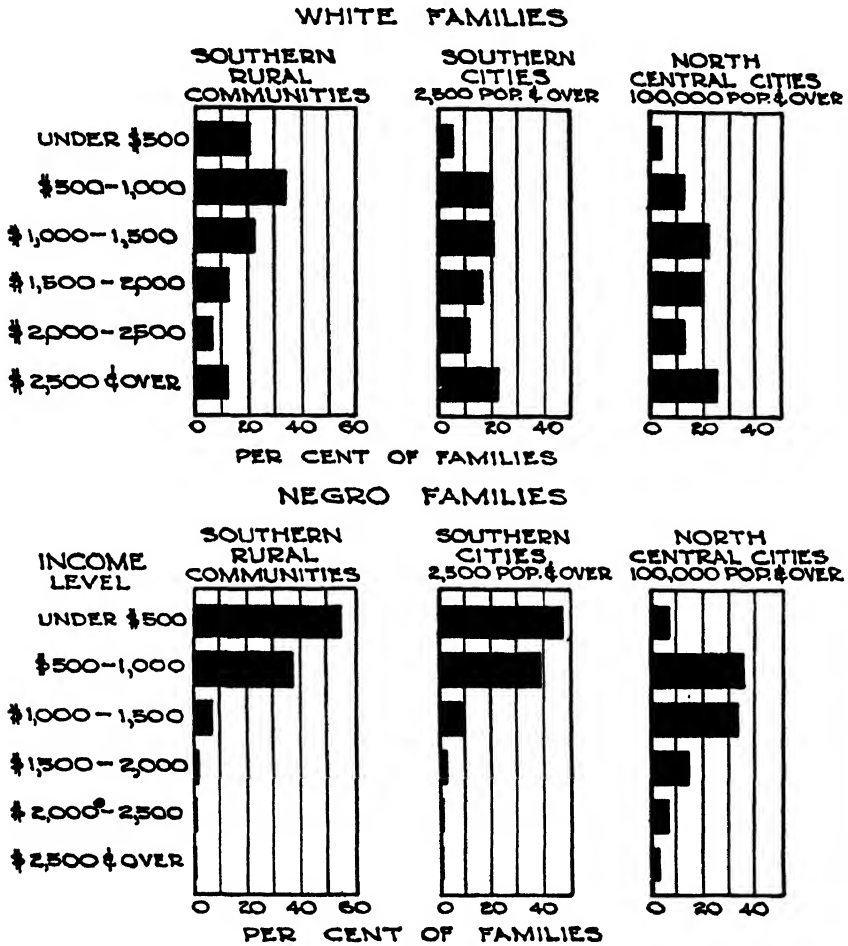


Fig. 35—5. Income distributions of white and Negro families (nonrelief) in three types of communities, 1935-1936.*

* Based on data from *Consumer Incomes in the United States*, Table 14, p. 28.

not ordinarily receive much lower wages than men for doing the same work, though numerous cases of this sort exist.

Negroes receive much lower incomes than do white persons. A comparison of incomes received by Negroes and whites is presented graphically in Figure 35—5. The explanation of the striking difference shown is very complex. Superficially it is related to differences in occupation, but actually it lies much deeper than that, in our entire social structure.

Income redistribution

The foregoing figures give an erroneous impression, for two major reasons: (1) The prevalence of low income families is underestimated because the figures are for money incomes received by nonrelief families only. (2) On the other hand, inequality is exaggerated because these figures fail to take into account the redistribution of income associated with the provision of government-financed goods and services to consumers. Common examples are highways, schools, the courts, parks, the Army and Navy. Such services are usually provided free to the consumer, except insofar as he pays for them through taxation or loans to the government. They are used by virtually all consumers, but the burden of financing falls somewhat more heavily on the high income groups than on the low income groups. Thus the lowest income groups probably receive substantially more real income from government services than they pay for, while the very high income groups pay for substantially more than they receive.

Relation Between Income Distribution and Resource Allocation

The economic system is an integrated whole, even when it is functioning very badly. The ways in which the allocation of resources is determined cannot be separated from income distribution. Any given distribution of income, on the one hand, will affect the way in which resources will be allocated. On the other hand, the processes by which resources are allocated are also the processes by which people receive their incomes.

Income distribution as the determinant of dollar votes

Resources are allocated largely on a *one-dollar one-vote* basis. The more dollars offered for a given product, the more of it tends to be produced, and whether the dollar is offered (spent) by a rich man or a poor man makes no difference. Thus, the person spending \$100,000 has a hundred times as much "voting" power over resource allocation as a person spending \$1,000. Such an unequal income distribution as exists in the United States naturally means, therefore, a very different allocation of resources from what would be obtained with greater income equality. If high income groups are willing to pay enough for imported wines and luxurious penthouses to give producers a profit, wines will be imported and penthouses will be produced instead of frankfurters and small homes or modest apartments. This will be the case in spite of malnutrition and unthinkable housing conditions among low income groups in both urban and rural areas.

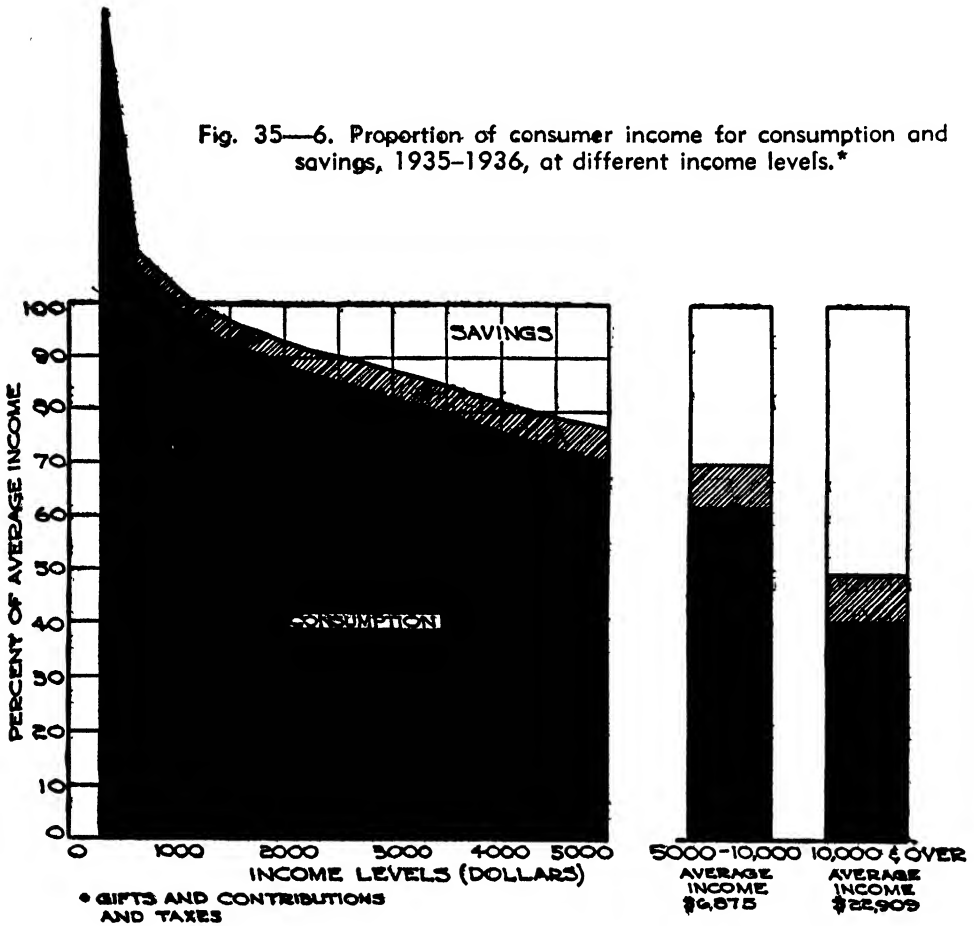
Expenditure patterns by income groups

It is worth while, therefore, to indicate briefly the expenditure patterns of different income groups, as an explanation of the actual "dollar

votes" cast on the allocation of resources in our society.⁸ These votes are cast for (1) future versus present consumption opportunities, and (2) various current consumption goods and services.

1. Votes for diversion of resources to future uses are made when money is saved and offered for investment. Figure 35—6 gives the estimated

Fig. 35—6. Proportion of consumer income for consumption and savings, 1935-1936, at different income levels.*



* Copied from *Structure of the American Economy*, Chart III, p. 10. Based on data from *Consumer Expenditures in the United States*.

average outlay of consumer units at different income levels for current consumption, gifts and personal taxes,⁹ and savings in 1935-36. The difference in the use made of income at the different income levels is very striking. The high income groups save a very large percentage of their incomes, and have to give up a good deal more in the form of personal (direct) taxes, so that much less than half the incomes of the

⁸ These estimates are based on *Consumer Expenditures in the United States*, by the National Resources Committee (U. S. Government Printing Office, Washington, 1939), a study in connection with the income study for 1935-36 already cited.

⁹ Income taxes, poll taxes and some personal property taxes.

\$20,000 and over groups goes for current consumption. As we move down the income scale, the percentage of income saved steadily falls and that spent on consumption steadily rises. Below the \$1,250 income level, not only are there no net savings but actually negative net savings; that is, persons and families in these lower income brackets actually spent more than their current incomes—a variety of deficit financing facilitated by borrowing, by buying on credit, by use of income received as gifts and charity, and so on. This phenomenon becomes more striking still if we stop to realize that the income groups which showed negative net savings in 1935–36 included about 60 per cent of the population. For the lower third of the income scale negative savings become very substantial. Figure 35—6 illustrates these facts very strikingly. Not until the \$1,500 level is reached do net savings become an appreciable sum in total or in percentage of income, and the group which has negative savings includes almost half the population.

While the savings data for 1935–36 may give us some clue as to the savings of consumers at different income levels in other years, we would make a serious mistake to apply them too generally. How much typical individuals and families will save clearly depends on many things besides incomes. If incomes drop suddenly people may be slow in adjusting their levels of living downward with their incomes, and vice versa. Moreover, savings from given incomes will be much greater in a year in which consumers are well stocked with durable goods than in a year that starts with these stocks very much depleted.

2. A brief survey of consumer expenditure patterns at different income levels has already been presented in Tables 1—1 and 1—2 and Figures 1—1 and 1—2, of Chapter 1.¹⁰ Turning back to these data, we see some interesting things. The greatest relative differences in expenditures between high and low income groups are for clothing, automobiles, recreation, and education; but food and housing are such important items in all budgets that the increase in actual number of dollars spent as income rises is greatest in these categories. In our society it would appear that “luxury” spending was spread pretty well through the entire budget. Though quantity consumption in food probably increases least, quality changes account for tremendous increases in expenditure as income rises. No figures will reveal these quality differences and hence the most interesting facts with regard to the effects of income on expenditures are buried in the details within each major category.

Effects of unequal income distribution on the allocation of resources

How would a modification of the present inequalities of income distribution in the United States affect consumers’ dollar votes? Some of the answers to this question are very obvious; others are more difficult to determine.

¹⁰ These data are for families. Figures for all consumers are available for 1941, but these data had not been broken down to separate families from single individuals when this book went to press.

First, it is necessary to consider the extent to which inequality in the distribution of income affects peoples' scales of preferences. The spending habits of high income groups may serve as examples and thereby exert pressures on those with lower incomes. Thus it may be that our data on saving in the lower and middle income groups would be changed somewhat if inequality of incomes were much less, that the elimination of very high incomes and very low incomes would change the saving habits of those in the \$1,000 to \$2,000 range. Moreover, spending in such a way as to "show off" one's high income, frequently described as "conspicuous consumption," might be very much diminished in a society in which inequality of income was much less than it is in ours. These considerations are not measurable, but should be taken into account in reasoning from existing data on incomes, expenditures, and savings.

Probably one of the most striking effects of paring off incomes at the top would be the reduction in the total amount of savings by individuals and families, since so much of the present savings comes from the very high income groups. Almost certainly fewer "luxury" items purchased only by the very high income groups would be produced—fewer resources would be allocated to producing yachts, high-priced automobiles, or extravagant hotels. If at the same time incomes at the bottom were raised the small cars such as the Austin, so common in Europe, might gain in importance here. There might be an increased demand for low-cost housing, a decreased demand for elaborate mansions. The least costly forms of recreation would probably gain relative to the more costly. There would probably be an increased use of resources in the production of milk, fresh fruits and vegetables, beef, some of the other nutritional foods as substitutes for wheat, corn, potatoes, and pork. Few commodities would escape the effects of a drastic income redistribution.

The effects of consumer spending on the distribution of income

While the distribution of incomes affects consumer spending directly, it is also *affected by* consumer spending. How much income a man will receive depends on the *functions* he and the agents he controls perform in the production of goods and services, and *on how much consumers are willing to pay* for these particular goods and services, in the production of which he is involved. Over long periods of time people may shift their investments, or their occupations. In the short run their incomes depend on current shifts in consumer demands for the things they are able to produce. Thus in the long run consumer spending influences the occupations into which people move, and the investments they make. In the short run consumer spending has a more limited effect on this allocation of resources, a more exaggerated effect on the *functional* distribution of income. These relations may be summarized tersely somewhat in this way: Personal distribution influences consumer dollar votes, consumer dollar votes influence functional distribution, functional distribution determines personal distribution.

CHAPTER 36

The Sources of Personal Money Incomes

WHY does Tom Smith have an annual income of \$1,500, Ed Green an income of \$5,000, and Jim Williams an income of \$100,000? What are the sources from which these incomes are paid? And what determines the quantities of income payments from each source?

The Flow of Money Through the Firm

In a very simple economy, a large portion of the revenues received in the sale of products become directly the personal incomes of the sellers of the goods. In a complex exchange economy with the high degree of specialization characteristic in America today, this is no longer true. The incomes of the majority of individuals in the United States are paid to them by firms. Some of the funds flowing through firms go directly to individuals as personal incomes, and are spent on consumer goods or saved by these individuals (and their families) according to their needs, preferences, and opportunities. But a large portion of the funds flowing through most firms goes to other firms; these sums become personal incomes only as subsequent firms may pay them out to individuals.

The path of a hypothetical dollar from consumer spending to consumer incomes •

The flow of money through firms from consumer spenders to ultimate receivers of personal incomes is a fascinating process. It would be almost impossible to follow completely the intricate path of a dollar spent initially by a consumer in a retail store, but by presenting a very much oversimplified model of what happens, we may get some understanding of the character of these money flows.

Let us suppose that John pays a dollar for a cotton tie purchased in a local haberdashery. What will happen to this dollar as it moves through the economy? Its travels may be marked out by a series of steps such as is given below. Each starred item indicates money received as consumer income once more. (Taxes and other government activities are ignored.)

Step 1. John spends the *dollar*, which becomes the revenue of the retailer.

* a. *Fifteen cents* of this dollar is paid out in personal incomes to clerks

in the store, to those who have invested in the enterprise, to the entrepreneur.

- b. *Eighty-five cents* is paid to the wholesaler, from whom another tie is purchased (in order to replace stock on the retailer's shelves).

Step 2. The wholesaler receives *eighty-five cents*.

- * a. *Ten cents* is paid out in personal incomes to those connected with the wholesale enterprise (workers, investors, and so on).
- b. *Seventy-five cents* is paid to a tie manufacturer.

Step 3. The tie manufacturer receives *seventy-five cents*.

- * a. *Forty-five cents* is paid out in personal incomes to those connected with the manufacturing enterprise.
- b. *Eleven cents* is paid out to producers of raw cotton.
- c. *Seven and a half cents* is paid out to coal and iron mining firms.
- d. *Seven and a half cents* is paid out to contractors for plant construction and repair.
- e. *Four cents* is paid out to utilities for light and power.

Step 4. (1) *Eleven cents* is received by producers of raw cotton.

- * a. *Seven and a half cents* is paid out in personal incomes to those connected with the cotton-growing enterprises.
- b. *Three and a half cents* is paid to other firms (for fertilizer, agricultural implements, and so on).

Step 4. (2) *Seven and a half cents* is received by mining companies.

- * a. *Four cents* is paid out in personal incomes to those connected with the mining enterprises (miners, investors, and so on).
- b. *Three and a half cents* is paid out to other firms (especially utilities).

Step 4. (3) *Seven and a half cents* is received by contractors.

- * a. *Three and seven-tenths cents* is paid out in personal incomes to those connected with the construction enterprise.
- b. *Three and eight-tenths cents* is paid out to other firms.

Step 4. (4) *Four cents* is received by utilities.

- * a. *Two cents* is paid out in personal incomes to those connected with the construction enterprise.
- b. *Two cents* is paid out to other firms.

By the time the dollar that John spent on a tie has been traced through all this labyrinth of exchanges, most of it has once more appeared as someone's personal income. There is still ten and eight-tenths cents, however, which is the gross revenues of firms that will further divide their payments in miscellaneous ways. Eventually the entire dollar (and no more) becomes consumers' incomes. This simplified picture of the travels of a dollar is represented diagrammatically in Figure 36—1. The width of the bands indicates the number of cents in each case. Black bands are flows from firm to firm: gray bands with arrows are flows from firms to individuals as consumer incomes.

That the picture here presented is extremely simplified as compared with what would actually happen can be quickly seen. For example, there is only one manufacturing firm indicated; such a firm would have to be extremely versatile, producing its own machines and office equipment, and even producing the steel with which to manufacture the machines, as well as manufacturing the final consumer product sold to the wholesaler. But it nevertheless presents an essentially accurate picture of the nature of the flow of money through firms in the American

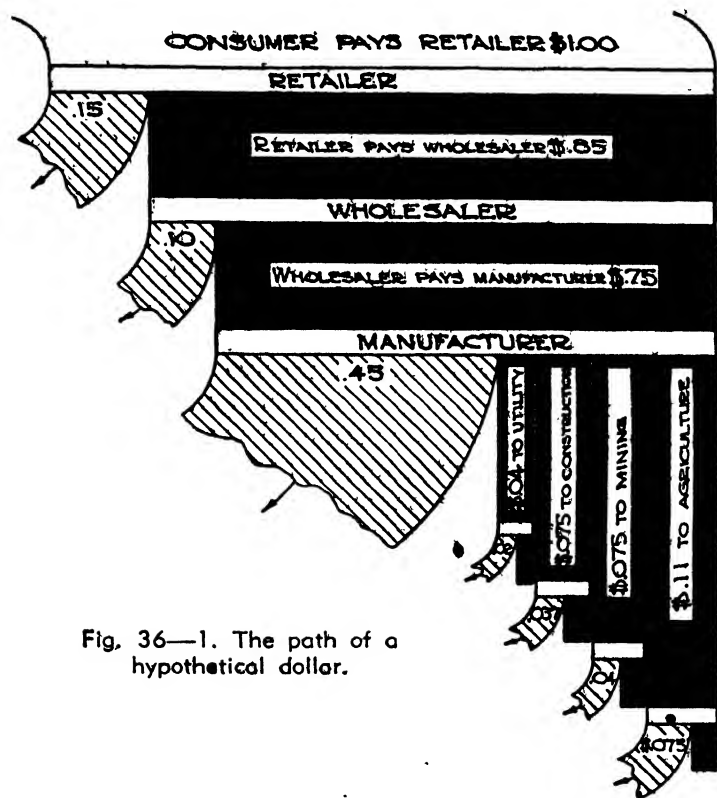


Fig. 36—1. The path of a hypothetical dollar.

economy, with the gradual syphoning off to individuals of incomes that again join the stream of consumer spending.

The functional division of money flows through the firm

Incomes received through a firm may be classified according to the basis on which they are obtained. The traditional classification into four groups is convenient: (1) incomes received as returns for labor services (including entrepreneurial wages), (2) incomes received as payments for services of material agents of production, (3) incomes received as returns in interest paid for the use of "loanable funds," and (4) incomes received as profits.

1. Payments for labor services make up the major part of the flows of money to individuals as personal incomes. The retail clerk and the manager of the retail store, the wage worker in the textile factory and his boss, the telephone operator, the carpenter, the farmer and the hired man—all of these people receive money incomes as payments for the labor services that they provide. Most of these payments in return for labor services are explicit costs, definitely agreed upon by employer and employee; but an important part of them are (the implicitly expected) returns for the labor services of men in entrepreneurial positions, as the farmer, or the proprietor of an independent grocery store.

2. Payments to owners for the services of material agents are often payments from one firm to another. The buying or hiring of machines, and the purchase of raw materials are clear-cut examples. The most common examples of payments by firms to individual owners of material agents are probably rentals of land and buildings; but these payments make up a small proportion of the total flow of funds through firms to owners of material agents. When a cotton textile manufacturer buys a loom he pays a firm manufacturing looms; only indirectly does this payment become personal incomes, as the money thus received is paid out to individuals by the loom manufacturer and by other firms (still farther back in the production process) that receive payments from the loom manufacturer. Similarly, when a cotton textile manufacturer buys cotton from a dealer his payments become the revenue of another business firm, which in turn disperses the money partly to individuals and partly to other firms. By the time payments for material agents have been syphoned back to individual income receivers through the chain of firms engaged in successive productive activities, these incomes appear in changed form—as wages, interest, and profits. Even payment made directly by firms to individual owners of land or buildings may be expressed, as will be explained later, as profits and “interest” payments on the funds invested by owners in the land or buildings rented.

3. Payment of interest for the privilege of using “loanable funds” is a universal part of business enterprise.

“Interest” as the term is used in common parlance is the extra amount that is paid back to a lender by a borrower of funds when he repays his loan. The economist points out that when a man uses his own funds instead of borrowing he must “implicitly” expect to receive back something above the sum with which he starts; otherwise he would lend his money out and obtain “explicit” interest from somebody else. “Interest” therefore may be defined as the price paid for the use of a given amount of loanable funds over some given period of time, whether these funds are in fact loaned from one person to another or not. Interest is most frequently described as an interest “rate,” as the number of dollars that will be paid for the use of each \$100 for some given unit of time, usually a year. Thus when an interest “rate” is quoted as 4 per cent per annum, this means that the payment for the use of \$100 for one year is \$4. Thus

far, interest has been only briefly mentioned as a production cost; it will be examined further in subsequent sections of this chapter.

Interest payments are received as personal incomes by private investors in corporate securities, that is, in stocks and bonds. They are also received by individuals lending money to noncorporate firms (single proprietorships and partnerships), and by owner investors in their own enterprises. But firms as well as individuals lend out funds, and these loans bring in interest payments as revenues of the lending firms; this is the special function of banking enterprises, and is a common practice of many corporations. Interest payments thus move both from firm to firm and from firm to individual.

4. Profits are a residual income, what is "left over." Where returns from a business are sufficient to pay all its creditors (including hired laborers, firms from which it has purchased, and individuals and firms from which funds have been borrowed) and to cover implicit interest and entrepreneurial wages, still leaving an excess over these amounts, that excess is positive profits. If all these costs are not fully covered the deficit is negative profits (loss).¹ Positive profits become personal incomes when they are paid out to owner-investors in an enterprise—to individual stockholders in a corporation, to partners or proprietors who have invested their own savings in their businesses. They are also paid from one firm to another when a firm holds stock in another enterprise; this is most strikingly illustrated in the case of corporate holding companies.

Plan of study

In tracing the path of a dollar from consumer spending through firms to consumer incomes, and in examining the functional division of money flows through firms we have pointed up the nature of the sources of personal incomes. These incomes are received partly in wages, partly in returns on investments of funds in the employment of productive services of all kinds. Factors determining wage incomes have been examined in considerable detail in previous chapters. In order to fill out the picture it is now necessary to look more fully into factors determining personal "property incomes."

All incomes from sources other than payments for services of labor are frequently called "property income." To own "property" is to possess a claim on income (whether the income is realized or merely potential). Property may take the form of stocks or bonds in a corporation; ownership of these pieces of paper is ownership of a claim on the income of the corporation. It may take the form of a mortgage on a farm; ownership of the mortgage is ownership of a claim on the income of the farmer, and if he fails to pay, on the income from the land mort-

¹ The difference between economic profits and accounting profits as reported by corporations was explained in Chapter 7.

542 THE SOURCES OF PERSONAL MONEY INCOMES

gaged. It may take the form of a deed to land, which involves a direct claim on the income from that land (and hence control over how the land will be used, whether for consumption purposes of the owner, yielding real income, or for productive purposes to yield him a money income). It may take the form of a patent right, which gives the owner claim on incomes obtained from exploitation of the patent, and hence again control over the way in which it will be exploited. The list might be extended considerably.

In the following sections of this chapter we shall proceed as follows: First, we shall examine property income in a hypothetical economy assumed to be in purely competitive general equilibrium; this enables us to simplify the initial step of our analysis. Second, we shall consider "pure profits"; such profits are an aspect of property income arising because of uncertainties and the immobility of resources in economic life. Third, we shall present a brief summary of monopolistic and monopsonistic elements in the distribution of income, including both wage and property aspects of income distribution. Finally, we shall employ the analytical concepts developed here and in preceding chapters to interpret some statistical data on sources of income.

Property Income and Purely Competitive General Equilibrium

If an economy were purely competitive throughout and were in a position of general equilibrium, profits of all kinds would be eliminated. In studying such a hypothetical economy, we may therefore focus on the other elements in the determination of property incomes. All property income accruing to individuals in such an economy may be ultimately classified as "interest" on invested funds. But the ultimate source of income is in the productivity of the physical agents of production (including labor) in which the funds are invested. We are now ready to trace through these sources of property income, examining their interrelations.

Net income from material agents of production

At any given time there exists in an economy a limited quantity of material resources that may be used in production. Consumers demand and are willing to pay for the goods in the production of which these material agents make a contribution. There is therefore a "derived" demand for the agents themselves, and incomes can be obtained from using these agents or selling or renting them to others who will use them. This derived demand is expressed through the sale values of marginal products of various quantities of each particular kind of productive agent.

Some material resources are very durable and are not readily depleted with use; others wear out quickly. A continuing income flow can only be obtained if resources that wear out are regularly replaced. This fact is easily illustrated. Suppose a man owns some horses that he rents out to those wishing to ride them. If he fails to replace the horses as they grow

old, or become disabled, he will find that before long his source of revenues disappears; he will have been "living on his capital." Net income from his property in horses is that income that he obtains over and above replacement of the horses. It is only that *net income* that would generally be regarded as the personal income of the man running the stable. This fact is recognized in defining income under the personal income tax. Similarly the United Shoe Machinery Company rents out machines for the making of shoes; if it fails to replace these machines as they wear out, its source of revenues will dry up. Allowances for depreciation in a business enterprise are recognitions of this important distinction between gross and net income. On the other hand, urban land sites do not wear out with use; they are still there, and other things equal they are still worth the same amount at the end as at the beginning of a year. The entire income obtained from such an "indestructible" resource is a net income to the owner; he needs no depreciation fund to keep his "capital" intact.

There are some kinds of material resources that wear out or are exhausted but cannot be replaced. Mineral resources are a clear example. In this case the replacement of the "capital" of the mine owner requires the setting aside of funds equal to the depreciation of the mine each year; these funds may then be invested in some other type of property claim. By such reinvestment elsewhere, the individual keeps his capital intact; but the society has used up the mineral resources. Replacement from the social point of view will have been accomplished only if resources of equivalent value have been "produced" to replace the depleted mineral resources. The net income from the mine is income over and above this replacement allowance.

This analysis suggests a convenient simplified classification of material agents into four types, which are approximated in various degrees by any given concrete case: (1) First, "reproducible-destructible" agents are physical agents of production that can be duplicated by productive activity of other agents, and that wear out or get used up. Machines of all kinds are an example. A part of the revenue from such agents is merely replacement of the value of the resources as they are used up; these replacement funds may be invested in production to "reproduce" the depleted agents, or they may be shifted into producing something else instead. Net income from such agents is the excess of revenue yield over such replacement allowance. (2) Second, "nonreproducible-destructible" agents are physical agents of production that, like mines or oil wells, are used up but cannot be duplicated by economic activity. A part of the income of such productive agents is again merely replacement of the value of the resources as they are used up; net income is the excess of revenues received over such replacement allowance. These replacement funds are necessarily used in production of *other* kinds of resources.² (3) Third, "reproducible-indestructible" agents are agents that can be duplicated by

² Unless they are withdrawn for consumption uses.

productive activity of other agents, but that are not depleted or used up. In some cases this is approximated by urban land that can be duplicated by the creation of filled-in land (the filled-in section of the Charles River in Boston is a notable example) but that does not wear out once it has come into existence. Since it does not wear out, such an agent yields a revenue which is in its entirety a "net" income. (4) Fourth, "non-reproducible-indestructible" agents are agents that cannot be duplicated by the productive activity of other agents, and that do not wear out or get used up. This category is again most nearly approximated by urban land, but by urban land in interior areas where there are no possibilities of increasing the quantity by creating new filled-in areas. The entire income of such a productive agent (assuming no taxes) is again a "net income."

"Rent" and "interest" in purely competitive general equilibrium

"Rent" as the term is used in ordinary conversation means the gross revenue received for the services of a fairly durable agent of production that is used by someone other than the owner. But the gross revenue obtainable from the agent is the same whether it is hired or owned by the user. In economics, therefore, one of the meanings sometimes attached to the term "rent" is the gross revenue received per unit of the agent, whether the owner hires it out or uses it himself. Under purely competitive conditions this rent would equal the sale value of the marginal product of units of this agent in each of its employments.³

If an agent were completely indestructible, this rent would be a "net" income. But if the agent wears out with use, the gross revenue yielded over any given period of time would be greater than the "net" income. It is therefore important to distinguish between what may be termed the "gross rent" yielded by a material productive agent, and the "net rent" remaining after the deductions for replacement of depleted value. The more quickly an agent is used up, the greater the proportion of gross rent that represents replacement only, and hence the smaller the proportion that will be net rent from the services of the agent.

In a hypothetical economy assumed to be in general purely competitive equilibrium, all productive agents would be employed where their returns were at a maximum. Investments everywhere, in all industries, would be earning the same rate of return—otherwise there would be shifts of investment from less to more lucrative channels. What would be the relation between "rent" and "interest" under these circumstances?

Investors in productive resources that are approximately nonreproducible and indestructible receive a revenue from those resources that is in its entirety a net return on the investment. Let us say that on \$1,000 invested in a piece of urban land there is a return of \$50 per year. If there is general purely competitive equilibrium, all other resources valued at \$1,000 would also bring in rents that net \$50 a year after

³ See Part III.

allowance for replacement. That is, net rents everywhere must be \$50 a year on property worth \$1,000. If a nonreproducible resource that had sold at \$1,000 had yielded an annual net rent over \$50, investors would have competed to buy that resource and its price would have been bid up. If it was reproducible, more of it would have been produced until a unit earned an annual net rent of only 5 per cent on the investment involved in its production (\$50 on every \$1,000 invested). If a resource that had sold at \$1,000 had yielded a net rent under \$50, investors would have turned to other things. Suppose for example this net rent were \$40. Either the price of the resource would have fallen to \$800, so that it would earn an annual return of 5 per cent; or, if it was a destructible resource, it would not have been replaced as it wore out and as a result the gross rent per unit remaining would have risen until it provided a 5 per cent return on replacement value.⁴

Thus annual "net rents" derived from all kinds of productive agents would, under conditions of general equilibrium, be in the same proportion to the sale value of the agents. The ratio of annual net rent to the sum invested would be just what would be paid as a rate of interest for the use of loanable funds with which such resources might be purchased.

The income that an investor in material agents received could be termed either "net rent" or "interest," since through the processes of equilibrium adjustments they would tend to become equal. For resources that were destructible this figure would be less than the gross rent received from the agent in any given period of time; it would be the net income after deducting depreciation allowance from gross rent.

"Wages" and "interest" in purely competitive general equilibrium

It is now time to ask what connection, if any, exists between wages and interest. The answer to this question hinges on the fact that a period of time elapses between investment in labor services and the obtaining of the revenue from the added product resulting. Suppose, for example, that a period of one year elapses between the payment of wages and the receipt of income from the sale of the product added. Those in possession of loanable funds will compare the advantages of using these funds to hire more laborers or more physical productive agents. If \$1,000 invested now in an urban land site will bring back its original sale value at the end of a year plus \$50 net income, then investment in the hiring of labor services will seem worth while only if \$1,000 spent on pay roll now will bring back at the end of the year \$1,000 replacement of the investment plus at least \$50 net income. By putting his money into these labor services, the investor is foregoing a \$50 net return that he could otherwise obtain elsewhere. The addition to cost resulting from the additional labor services is in fact not only the \$1,000 of pay roll, but in addition the \$50 interest for the use of loanable funds for one year.

⁴ This increase in the yield per unit of the agent would of course be due to diminished outputs and hence higher prices of the products turned out by the diminished quantity of the resource remaining.

In previous chapters it was concluded that if an economy were in purely competitive general equilibrium, each worker would be receiving for his services a sum equal to the sale value of the marginal product that he contributes. This statement must now be modified. He will be receiving as wages a sum equal to the sale value of the marginal product resulting from hiring him *minus* interest on that sum for the period of time between paying his wage and recouping the resulting revenues. Although interest received from such investments appears to be deducted from the sale value of the marginal product contributed by the worker, it is nevertheless classified with other sources of interest as "property" income. The interest payments are received by individuals in possession of property in the form of claims on returns for the use of loanable funds; they are not received as payments for the labor services of the recipients.

The rate of interest and the allocation of resources

The principal conclusions of the preceding sections, as they throw light on the rate of interest, may be summarized in one brief sentence. In an economy characterized by pure competition throughout, adjustments would lead to the same rate of return on all investments. This rate of return is the equilibrium rate of interest.⁵ There would be no advantage to anyone in changing.

Preceding discussions serve to point up the significance of interest rates in their influence on resource allocation. When net income is stated as a proportionate return on investment, comparisons between different investment opportunities are facilitated; and since investments can take so many thousands of different forms, it is only through such comparisons of rates of return that informed decisions can be made. Would it be better to make several investments of \$100,000 each in establishments making woolen goods out of raw wool, spindles, looms, and mill labor, or to put \$500,000 into a factory to produce washing machines? The net income in one of the woolen factories might be much less than that in the washing-machine factory and yet it might be a more profitable investment if the net return per dollar of costs is greater. On the other hand, it may be that though it takes more money to establish and operate an efficient washing-machine factory the net return per dollar of costs is greater there than in the wool business. The decision to invest in one or the other will depend not on a direct comparison of costs, or even of net income; it will depend on the net returns per dollar expended in one line or the other.

A decision to invest in wool production rather than the manufacture of washing machines means that funds to purchase the services of productive resources will flow through firms in the woolen industry, and resources will tend to be channeled into that industry rather than into the production of

⁵ The literature of economics is rife with refined and intricate disputes over the exact nature of factors determining the level of the rate of interest. Most of these arguments turn on the kind of economy that is hypothesized for the purposes of the experiment in logic. There is no need to go into this maze of elaborate analysis here; indeed to do so would be futile.

washing machines. Steel will be used to produce looms and spindles instead of machinery to manufacture washing machines. Men will be employed as weavers and spinners instead of inspectors or workers on the assembly lines in the washing-machine factories. When the rate of return on investment in the woolen industry is greater than the rate of return on investment in making washing machines, this must mean (assuming a purely competitive economy) that consumers are willing to pay more for the product of \$100 worth of resources used in making woolen goods than they would pay per \$100 worth of resources used to make washing machines. By comparing rates of return in different investment opportunities, and by choosing those in which anticipated returns are highest, businessmen are thus responding to consumer preferences as to how resources should be allocated.

"Pure Profits"

Profits are the excess (if there is any) of the gross revenues of a firm over costs in any given time period. If revenues are less than costs, profits are negative—usually designated as "losses." These "costs" include the promised or anticipated incomes that were necessary to attract into the firm those quantities of various physical productive agents employed, plus the "going rate of interest" on invested funds. They include all "opportunity costs" whether explicit or implicit. Since profits are merely any excess over these costs, they are a "residual" income—what is "left over."

A positive residual income above opportunity costs may arise under either or both of two principal kinds of situations: (1) In the dynamic changes of economic life there may be unanticipated developments that bring excess returns to firms fortunately placed until there has been time for firms and resources to shift in response to these profit opportunities and thus to eliminate them. (2) Entrepreneurs in a particular industry may have been able to establish themselves in a position that excludes newcomers, enabling the insiders to obtain profits through restrictions on production. Profits of the second type would not appear where there was freedom of entry to an industry; they are not the subject under discussion in this section. It is profits of the first type only that are designated as "pure profits."

The "profit motive" and elimination of "pure profits" in long-run adjustments

The American economy is often described as a "private-profit economy." Actually, as previous analysis has shown, pure profits tend to be eliminated in the long run wherever there is freedom for new firms to enter the field. When new firms enter a purely competitive industry in response to profit opportunities, the result is the lowering of the price of the product and the raising of the prices of productive services until profits are wiped out. When firms enter in hitherto oligopoly or oligopsony situations,

profits again tend to be eliminated, this time not only through price effects in selling and buying, but also through a dividing of production into smaller and smaller parcels so that each firm is operating at a less and less efficient level. If profits tend to be eliminated where there is freedom of entry, is it then meaningful to characterize the economy as a "private-profit system"?

The answer is most assuredly, yes, even when profits are defined narrowly as pure profits only. Anticipated profits and losses are motivating forces that bring about readjustments in the allocation of resources to different industries. Anticipated profits attract new firms and resources into an industry; anticipated losses lead old firms and resources to be withdrawn instead of replaced. By this very process pure profits and losses tend to be eliminated. But in the dynamic environment of the real world they continuously crop up again, exerting new pressures to shift productive agents and their services from old to new uses. They arise not merely because there is change, but because *(1) change and the results of change are in considerable part unpredictable and cannot be accurately anticipated in the plans of entrepreneurs, and (2) productive services cannot be immediately shifted from one use to another—because such shifting takes time.*

The major kinds of changes giving rise to such profits and losses may be classified as follows:

1. Widespread fluctuations in general business activity
2. Changes in demand for one product relative to others
3. Changes in the supplies of productive services to an industry
4. Innovations

The effects of each of these kinds of change is considered briefly in the following paragraphs.

Widespread fluctuations in general business activity

Most spectacular and very important are the widespread fluctuations in general business activity from "prosperity" to "depression" and back again. During periods of prosperity widespread profits arise, largely from forces outside the direct control of the individual entrepreneur. During periods of depression widespread losses occur in a similar manner. The phenomenon is in considerable part one of differential price level changes, since typically selling prices have risen faster than costs when prices are moving up, and have fallen faster than costs when prices are moving down. To the extent that prices of productive services are not bid up rapidly enough to keep in line with rising product prices, pure profits result. To the extent that prices of productive services fail to drop in line with falling product prices, negative profits (losses) occur.

General business fluctuations and price level changes are perhaps the most important types of change giving rise to pure profits and losses, since they are so widespread in their simultaneous effects. Further analysis of this situation will be undertaken in later chapters.

Changes in demand for one product relative to others

The demand for a product is seldom stable for long. Consumers spend more of their incomes now on one thing, now on another. Entrepreneurs are constantly faced with the problem of predicting what the future demands for various products will be; and this task of predicting is especially important since so much production is round-about and hence necessarily for a future market. When entrepreneurs fail to anticipate shifts in demand with sufficient accuracy, they may fail to enter in sufficient numbers into the production of goods the demands for which turn out to be great, and they may find that there is a relative excess of production of the goods the demands for which have dropped off. There will be a lag in adjustment to the change, with associated profits or losses during the interval of adjustment. That is, errors of anticipation combined with immobility of productive agents into (or out of) an industry will result in temporary "pure profits" (or losses). If the demand for a particular product increases more than most entrepreneurs expect, profits will occur until there is time for more firms to move in and get established with plant and equipment in operation. If demand falls below expectations, losses will occur until there is time for some firms, with the resources that they employ, to move out of the industry. These facts keep entrepreneurs trying to outdo each other in anticipating correctly the changing needs of society.

Changes in supplies of productive services to an industry

Changes in the supplies of particular productive services to an industry may reflect changes in the offerings of that kind of productive service throughout the economy. For example, new rich mineral deposits may have been discovered, increasing the supply of copper; or perhaps there has been a great immigration of common labor; perhaps previous sources of supply of a raw material, say rubber, have been cut off and supply has decreased. The supply of a productive service to a particular industry might also change owing to an increase or decrease in the bidding for this service by entrepreneurs in other industries; if it is wanted more than before in other industries, its supply to this industry will of course decrease, and vice versa. Other things being equal, an increase in the supply of a productive service to an industry will make it available to entrepreneurs in that industry at lower costs than previously, and a decrease in the supply to an industry will have the opposite effect.

The profit or loss effects of changes in the supply of a productive service to an industry will depend on whether the service on the one hand is repurchased or rehired at frequent intervals so that it is relatively variable, or on the other hand, is a relatively fixed agent replaced only at infrequent intervals. In almost every enterprise there are some costs, as in the purchase of machines, that are incurred only at infrequent intervals and that are expected to bring back revenues over fairly extended time periods

Increases in the supplies of such productive services offer advantages in costs only to those enterprises in a position to make new purchases of the agents involved. For this reason the analysis of the profit and loss effects of changes in supplies of productive services to an industry is necessarily more complex than is the analysis of the profit and loss effects of changing demands; and for this reason analysis here will be simplified by assuming pure competition. Removal of this assumption of pure competition would not alter conclusions in any essential respects so long as the assumption of freedom of entry, basic to this entire discussion of pure profits, was retained.

Let us suppose first that the supply of a relatively fixed agent, say a certain kind of machine, were to increase. New firms could make profits purchasing this machine under the new, more favorable circumstances, paying lower prices for it. As these new firms appeared, the price of the product of the industry would fall. The older firms would no longer cover their original costs, which included a higher price for the machine. Such firms would, however, replace old agents as contracts expired, paying the new lower rates for the replacements. During the interval of adjustment the old firms would have been operating at a loss, though profits were temporarily available to the new firms first appearing in the industry. If, instead of falling, the price of the machines had risen, old firms would not have suffered losses on their original investments; but when it came time for replacements some of them would probably have withdrawn from the industry.

Now let us suppose that there has been a change in the price of the services of a variable agent, say mill labor, instead of in the prices of the relatively fixed agents. What of the position of the established firm? Investments in relatively fixed agents such as looms must be based on expectations of prices of the other productive services that will be used in combination with the fixed agents. If after services of relatively fixed agents have been contracted for on the assumption of certain expected variable costs (such as wage rates), the variable costs change, this will mean too many or too few fixed resources in the firm and the industry. For example, if wages should rise in such a case (perhaps because of increased demand for this labor in another industry, or for some other unanticipated cause) too many entrepreneurs would have made investments in fixed agents in this industry. They would incur losses. When it came time for replacement of fixed agents some firms would withdraw until those remaining were able to cover their costs. If wages had dropped, on the other hand, profits would have appeared until such time as more firms could become established in the industry.

Innovations

The results of the introduction of new products or new productive methods are never completely predictable, and are often almost completely unpredictable. Naturally the entrepreneur who introduces such "innovations" anticipates that they will "pay" or he would not undertake

the change; but whether his expectation will be more or less than fulfilled he cannot accurately predict. It may be that either profits or losses may result from his innovation.

If an entrepreneur is fortunate with his new product or productive method, profits of innovation will appear and they will attract newcomers to the field unless entry is in some way restricted. Profits of innovation, like profits arising from unanticipated demand and cost changes, tend to be eliminated (unless entry is restricted) by the movement of resources into profitable areas. After a successful innovation, firms already committed to older productive methods or to products replaced by the innovation are likely to find themselves with losses until they may also adopt the innovation. For them the innovation has caused unpredictable changes in demand or in costs or in both.

"Risk" versus "uncertainty"

In discussing the importance of dynamic change in leading to pure profits and losses, much emphasis has properly been placed on the unpredictability of such changes as the crucial factor, coupled with lagging adjustment to changes. Some unpredictable events do not, however, give rise to profits and losses, even when adjustments are lagging. An individual entrepreneur cannot predict accurately whether his building will burn down within the next year, or whether his cash-box may be robbed, or whether hail will ruin his crops. Yet such individually unpredictable happenings need not lead to profits or losses, since the individual risk involved may be avoided merely by taking out insurance. The insurance premium is then a regular cost. The characteristic of "insurable risks" that makes them "insurable" is the possibility of predicting the average losses (as from fire, burglary, and so on) when many individuals are involved. The insurance company collects enough from all clients to pay the losses of the few. By covering a large number of cases, the cost becomes predictable to the insurance company, and the individual is able to convert his unpredictable risks into entirely predictable costs.

When risks are thus insured they do not give rise to pure profits or losses. It is the many uninsurable risks associated with dynamic change that are the ones that lead directly to pure profits and losses. These are often distinguished from insurable risks in economic terminology by calling them "uncertainties."

It is frequently argued that pure profits are not a "functional" share in distribution since they are not a necessary payment for the performance of a function in production. But the *possibility* of obtaining profits is in fact necessary in order to persuade people to perform an important function. Someone must take the chances associated with the uncertainties of economic life, and all of us perforce take some of these chances. The entrepreneur-owner of an enterprise, however, occupies a special role among investors. He takes the first chance of loss and he also has the right to profits. These profits and losses may just neutralize each other taking all firms together, so that profits may average zero. But the

possibility of gain nevertheless remains a necessary basis to attract men into investments where possibilities of loss are also great.

Monopolistic and Monopsonistic Elements in the Distribution of Incomes

In many important respects monopolistic and monopsonistic market situations affect the distribution of personal incomes. Most of these factors have been taken into account in analyses in earlier chapters, and others can be discussed in detail only after an examination of the money and banking system and general business fluctuations. Despite their great importance, they will therefore be considered only briefly here.

The wages of labor

1. Where labor groups form monopolistic organizations that exclude outsiders, the members may gain for themselves wage incomes greater than would otherwise be available to them. This is true whatever the position of entrepreneurs buying their labor services, provided the exclusions on entry to the occupation are really effective. The organized workers usually gain these higher incomes at the expense of other groups, especially of laborers who are excluded and of consumers of the ultimate product.

2. Where labor is sold purely competitively to entrepreneurs who are monopsonistic in buying, wages are likely to be depressed below the level they would otherwise reach. To a large extent such depression of wages probably diverts income from the workers to exclusive entrepreneur-owners in the form of profits associated with restrictions on entry.

3. Where monopolistic labor groups bargain with monopsonistic employer groups, the resulting wage level is probably determined largely by the degree of strength of those on each side of the bargain. That strength will depend, for example, on the attitudes of the general public and the relative political power of the labor and employer groups. It will also depend on the "staying power" of each in case of strike, that is, on how serious losses to the employer are, how great his financial resources are, and on how large the trade-union's "war chest" is. No general conclusion as to how results of such bargaining would compare with a purely competitive labor market can be reached; but frequently in such cases both entrepreneurs and laborers are able to hold up their incomes by excluding competitors from their particular spheres of activity.

Rates of interest

In presenting a theory of interest, it was assumed that the economy was in a position of purely competitive general equilibrium, and that interest rates were determined in such a free market. This analysis was of course very much oversimplified. Two of the most important omissions of the purely competitive analysis were these:

1. The quantity of loanable funds appearing in investment markets depends not only on the incomes of individuals and businesses and their

decisions to save and invest, but also on money and credit which is "manufactured" or "created" by the banks. This peculiar characteristic of the offerings of loanable funds is not in itself a monopolistic feature of the determination of interest rates; but it does help lay the basis for what are essentially centrally managed rather than purely competitively determined interest rate charges. This introduces a technical problem that will be explored in some detail in Chapter 39.

2. Where entrepreneurs in a particular industry are able to exclude outsiders they may block the tendency toward equalization of returns on investments in different sectors of economic activity. This is somewhat analogous to the differential returns to laborers of comparable skills according to whether they are inside or outside a monopolized occupation. Wherever the return on investment funds is greater than that which would be available in alternative opportunities in free investment areas, the excess is a kind of profit. The very concept of interest as a production cost rests on the concept of opportunity costs reflecting investment opportunities in free investment markets.

Profits of exclusive position

Pure profits resulting from unanticipated change with lags in adjustment may arise in any type of market situation. Such profits are in their very nature a temporary phenomenon. In addition to such "residual" incomes resulting from economic uncertainties, there is another factor which may give rise to profits. Wherever there are restrictions on the free entry of firms into a particular industry or production area, insiders may obtain over long periods returns in excess of production costs. These profits are not eliminated, because actions in response to such profits are blocked by exclusions of new firms and new investments. There are many terms by which economists distinguish profits gained through practices of exclusion; here they will be called "profits of exclusive position." The phrase is long and somewhat awkward; but it does describe the situation.

Firms in monopolistic positions maintain their controls over prices by excluding new enterprises from competing with them in the sale of a product. Firms in monopsonistic positions maintain these positions by similarly excluding new enterprises from competing with them in the purchase of a productive service. The "profits of exclusive position" that arise as a result of such practices mean returns to investors here that are greater than returns available in free market areas. Such returns remain high both because costs are not bid up to wipe out profits and because selling opportunities of the insiders are not cut by the force of new competition.

Some of the techniques of restriction of entry have been described in earlier chapters. A clearer understanding of the nature of profits of exclusive position may be attained by taking one or two of these techniques as examples and studying their effects.

First to suggest itself is the maintenance of an exclusion position through the control of a patent or a family of patents that are basic in

the production of a good. Many examples of such controls were cited in Chapter 26. The recent control of all important radio patents by a small group of dominant firms provides a clear-cut example. Through their joint ownership of patents, the dominant radio producers were able to maintain exclusive control over the production of radios and to hold up price by limiting production. There was nothing in this situation that would lead to a soaking up of the resulting profits; new firms could not enter to compete in selling, and there was no reason why the prices of productive services purchased by the exclusive group of firms should be bid up. A patent is of course worth to the firm or group of firms the profits of exclusive position that it makes possible; and for this reason such profits are sometimes called a "monopoly rent," a return arising from possession of the instrument by which the profits are made possible.

Another technique by which a monopolistic position may be maintained is the exclusive ownership of a crucial raw material, as exemplified by the Aluminum Company of America. As sole owner of the bauxite deposits that are the raw material of aluminum production, this corporation was able to obtain and to maintain an exclusive position in the production of aluminum ingots, and to hold up the price of aluminum. New firms could not enter the industry and compete in the sale of the product and thus cut down on sales revenues of the Aluminum Company. Moreover, there was no free market in which the prices of the bauxite could be determined. The cost of the bauxite deposits to the Aluminum Company was the initial purchase price; no repricing through competitive bidding was permitted to occur. The excess of returns that can be obtained by the corporation because of its exclusive control over the bauxite deposits is a measure of the worth of the deposits to the corporation; but this is more than the opportunity cost of attracting the bauxite into the firm. Insofar as the profits of the Aluminum Company of America are made possible by its control over the bauxite deposits, these profits might again be regarded as a "monopoly rent" of the instrument by which the exclusive position and the profits associated therewith are made possible.

Frequently monopolistic and monopsonistic positions are maintained through a combination of techniques and manipulative devices that cannot be so easily isolated as patents or ownership of crucial raw materials. But in such cases also there is a restriction on the entry of firms and on the flow of investments into the industry or production area; and these restrictions prevent the cutting down of sales opportunities of the original firms, and the bidding up of their costs. Such profits might also be regarded as a kind of "rent" of an exclusive position, though the instrument providing the profit possibilities is less clearly defined.

In this discussion it is apparent that "pure profits" and "profits of exclusive position" are very different kinds of things. Pure profits are temporary, and arise out of dynamic changes that are not fully anticipated and to which adjustments take time. Profits of exclusive position are returns in excess of costs that are made possible by the use of instruments or techniques that prevent the entry of competing firms or

investments; these profits may be maintained continuously over long periods of time.

Sources of Personal Money Incomes in the United States

Comparison of statistics of income sources with income concepts in economic analysis

In order intelligently to interpret statistics of income it is necessary to understand how they are related to the analytical framework that has been presented in preceding pages. A brief comparison between commonly used statistical categories and the concepts of economic analysis will therefore be undertaken here.

Statistics of income sources frequently separate them into five major classifications: (1) wages and salaries, (2) entrepreneurial withdrawals, (3) dividends, (4) interest, and (5) rents and royalties. The first of these is almost entirely income from labor.⁶ The others require more detailed comment.

Entrepreneurial withdrawals are the net incomes of entrepreneurs; and these incomes are in large part essentially "entrepreneurial wages"; hence they also represent labor incomes. A minor part of entrepreneurial withdrawals is returns on the investments of entrepreneurs in a business, made up of "economic interest," "pure profits," and "profits of exclusive position."

Dividends and interest are separated in these statistics of income sources, including under "interest" only payments of amounts agreed upon in contracts between debtors and creditors. The amount of interest stated in a contract is the *maximum* amount that the lender may receive.

On the average, investors receive back something less than this contract interest. Interest payments are sometimes in default and when bonds or mortgages fall due they are not always paid back in full. Consequently, contract rates must permit at the maximum some return over the opportunity cost that would be involved in attracting funds into investments where there was no chance of default. Included in contract interest, and hence in statistics of income appearing as "interest," is therefore an element of "pure profits," a payment for lending funds in the face of economic uncertainties. Only in the case of investments that are practically riskless, as government bonds, can contract interest be reasonably regarded as identical with "economic interest." Otherwise investors receiving the full amount of contract interest are in fact getting a return that is in part pure profit. Investors receiving much less than the interest payments contracted for may be getting back less than "economic interest"; but if they are receiving only a little less than the contractual obligation they may be getting full economic interest.

⁶ When corporations are closely held by a small group of people, salaries to owners and family members employed are frequently adjusted according to the fortunes of the business. Such salaries therefore frequently include an element of positive profits. They may also occasionally be cut to absorb losses.

"Dividends" are the payments received by the owners of stocks in corporations. That part of dividend payments the anticipation of which would be necessary to attract investment funds in the absence of uncertainty is economic interest. These are implicit production costs. If there is a positive residual it may be partly "pure profits," partly "profits of exclusive position." The possibility of obtaining positive pure profits is a necessary incentive to attract investment funds where there is economic uncertainty; and such positive profits tend to be balanced in the long run by the negative profits (losses) of those whose dividend receipts were insufficient to cover "economic interest." Profits of exclusive position, which are also included in dividend payments, cannot be separated statistically from interest and pure profits though they are very different in the nature of the factors that cause them to appear.

"Rents" as recorded in statistical reports are incomes paid to owners of physical property by those using it. Examples would be rental of urban houses by consumers, rental of farms by farm tenants, rental of store buildings by retailers.⁷ These figures are for "net" rental incomes; deductions for replacement of property depreciation have been made. Such deductions would, for example, cover replacement of housing as it depreciates, of soil fertility, and so on. The remaining net income from rentals is then a composite of different types of income sources as defined in economic analysis. It includes interest received on investments in the property offered for hire; it includes in part pure profits accruing to fortunate original investors in these rental properties where unanticipated changes have worked in their favor; it includes in some instances profits of exclusive position, illustrated by the operations of monopolistic "real estate rings" in some urban communities.

"Royalties" are payments of miscellaneous sorts—statistical data on royalties are usually confined to royalties paid for the privilege of extracting natural resources from the ground (as in the operation of a gravel pit or a copper mine).⁸ In many respects royalties are similar to rents as income sources. They are also net incomes after deductions for depreciation. Such deductions might be depreciation allowances covering depletion of mineral resources.⁹ The remaining net income from royalties is then, like net rents, a composite of economic interest and economic profits. Such income represents in part economic interest on investment in the quarry or in the copper mine. It represents in part pure profits arising from economic uncertainties that happen to have developed favorably to the investor in the quarry or in the mine. And it represents in part profits of exclusive position, these profits arising because of the possibility of excluding entry of new firms and new investment.

⁷ Estimates of the rental value of owner-occupied houses are commonly included in income estimates, though only cash rents are included in the data presented here.

⁸ Royalties are also paid for rights to exploitation of patents and other exclusive legally protected privileges.

⁹ Or allowances covering the costs incurred in research directed to the development of patented technological innovations.

Principal sources of income in the United States

Data on sources of personal incomes in the United States in the years 1919 to 1937 are given in Table 36—1 and presented graphically

TABLE 36—1 *
AMOUNT AND PROPORTION OF NATIONAL INCOME PAID OUT, 1919-1937
(In millions of dollars)

Years	Wages and salaries †	Entrepreneurial withdrawals ‡	Dividends	Interest	Rents, royalties and balance of international payments	Grand Total
1919	\$36,145	\$11,958	\$2,895	\$2,925	\$2,455	\$56,378
1920	42,667	13,838	3,215	3,279	2,767	65,766
1921	34,423	10,268	2,932	3,410	2,246	53,279
1922	36,659	10,224	3,006	3,535	3,497	56,921
1923	42,255	11,165	3,823	3,772	3,651	64,666
1924	42,494	11,356	3,762	3,997	3,917	65,526
1925	44,494	11,648	4,362	4,249	3,920	68,672
1926	46,985	11,804	4,736	4,410	3,655	71,590
1927	47,204	11,781	5,036	4,678	3,471	72,170
1928	48,717	11,940	5,362	4,976	3,591	74,586
1929	51,509	12,296	5,978	5,202	3,569	78,554
1930	47,551	11,581	5,801	5,393	2,965	73,291
1931	40,188	9,848	4,335	5,295	2,366	62,032
1932	31,563	6,887	2,745	5,019	1,811	49,025
1933	29,596	7,214	2,209	4,710	1,587	45,316
1934	34,051	8,021	2,793	4,862	1,783	51,510
1935	36,679	8,729	3,038	4,725	1,966	55,137
1936	41,906	9,565	4,284	4,652	2,179	62,586
1937	46,728	10,441	5,010	4,656	2,496	69,331

PERCENTAGE DISTRIBUTION

1919	64.1	21.2	5.1	5.2	4.4	100.0
1920	65.0	21.0	4.9	5.0	4.1	100.0
1921	64.6	19.3	5.5	6.4	4.2	100.0
1922	64.4	18.0	5.3	6.2	6.1	100.0
1923	65.4	17.3	5.9	5.8	5.6	100.0
1924	64.9	17.3	5.7	6.1	6.0	100.0
1925	64.7	17.0	6.4	6.2	5.7	100.0
1926	65.6	16.5	6.6	6.2	5.1	100.0
1927	65.4	16.3	7.0	6.5	4.8	100.0
1928	65.3	16.0	7.2	6.7	4.8	100.0
1929	65.6	15.7	7.6	6.6	4.5	100.0
1930	64.9	15.8	7.9	7.4	4.0	100.0
1931	64.8	15.9	7.0	8.5	3.8	100.0
1932	64.4	16.1	5.6	10.2	3.7	100.0
1933	65.3	15.9	4.9	10.4	3.5	100.0
1934	66.1	15.6	5.4	9.4	3.5	100.0
1935	66.5	15.8	5.5	8.6	3.6	100.0
1936	67.0	15.3	6.8	7.4	3.5	100.0
1937	67.4	15.1	7.2	6.7	3.6	100.0

* Source: *Structure of the American Economy* (National Resources Com.), June 1939, Table XIII, p. 380. Original Source: 1919-28, National Bureau of Economic Research; 1929-37, Dept. of Commerce, Bureau of Foreign and Domestic Commerce. National Bureau figures spliced onto Commerce figures upon basis of 1929 ratio.

† Includes work relief wages, employers' contribution to social security, and other labor income.

‡ National Bureau estimates adjusted to segregate entrepreneurial withdrawal from salaries and wages in service and miscellaneous industries on basis of average ratio of those items to total income. 1930-35 in Commerce estimates.

558 THE SOURCES OF PERSONAL MONEY INCOMES

in Figures 36—2 and 36—3. Data listed separately in the table as "interest," "dividends," and "rents, royalties, etc.," are combined in the charts under the general heading of "property income." Figure 36—2 shows the actual income derived over these years from wages and salaries, based on the same data but instead of giving the actual income figures it shows the percentage of total income from each general source over the period.

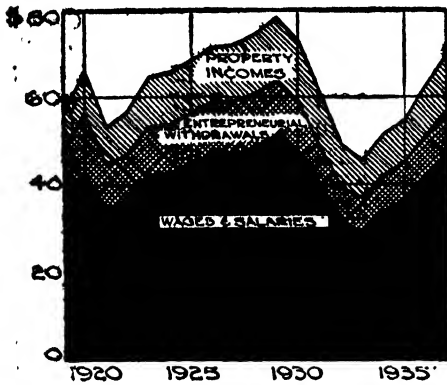


Fig. 36—2. Amount of national income paid out (in billions), 1919–1937.*

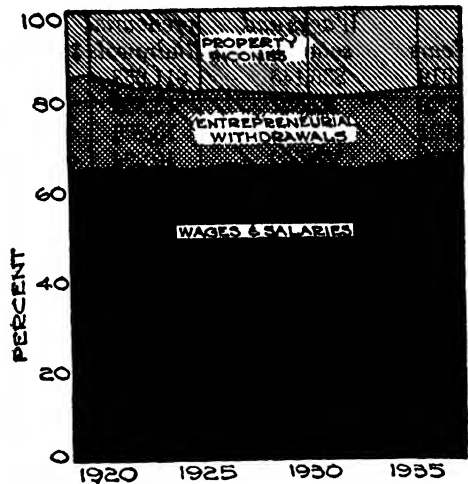


Fig. 36—3. Proportionate distribution of national income paid out, 1919–1937.*

* Reproduced from *The Structure of the American Economy*, p. 83.

Three facts stand out: (1) Most people receive their incomes primarily from wages or salaries, and this category has consistently accounted for almost two thirds of total national income; (2) there are extreme fluctuations in national income over even brief time periods, as is shown vividly in Figure 36—2; (3) there is a remarkable stability in relative shares of national income in each broad category; but here the charts hide some important differences. They fail to show that wages of common labor fluctuate much more violently than do salaries. And they show less variation in property income than in fact occurred, because corporations continue to pay dividends out of reserves in bad years while they pile up reserves in good years.

"Capital gains and losses"

An individual's net income is what he receives over and above maintenance of his "capital." This fact has already been discussed in connection with the defining of "net rent" in previous sections of this chapter. Any income over and above that necessary to replace the value of depreciating capital is a net income. This is a relatively simple concept

in the abstract; but it is not so easily arrived at in concrete situations as in analytical abstractions. The income accruing to an individual in a particular year may not be fully apparent; gains may be in part hidden in appreciated value of investments, and losses may be hidden in depreciated value not fully accounted for. This becomes a serious practical problem, for example, in estimating the base of an individual's personal income tax.

"Capital gains" may be roughly defined as increases in the sale value of a piece of property, whether this property takes the form of a direct claim on the ownership of particular real estate or other physical equipment, or a more generalized claim on the assets (hence income) of a corporation as evidenced in the ownership of stocks and bonds. Analogously, "capital losses" may be roughly defined as decreases in the sale value of a piece of property. Instead of estimates of these gains or losses being made year by year, they are commonly recorded as additions to or deductions from taxable income only at times of sale of property. They are of course not gains or losses of the particular year in which the sale takes place but rather of the entire interval between original purchase and present sale, and this is taken into account in income taxes. The gain or loss between purchase and sale is designated as "realized" capital gain, or loss. There are many refined technicalities concerning how such income should be reported for income tax purposes, but these are merely attempts to apply equitably the essential concept as outlined in these oversimplified statements. The figures on income sources in the preceding section ignored this item of the capital gains and losses per annum; but they are in fact extremely important in the higher income groups, and they change very markedly with changes in the total national income. Estimates of the per cent of total incomes in the form of realized capital gains and losses for the years 1926, 1929, 1932, and 1935 are presented in Table 36—2.

TABLE 36—2*

REALIZED CAPITAL GAINS AND LOSSES AS PER CENTS OF TOTAL INCOME

<i>Year</i>	<i>All income classes</i>	<i>Incomes under \$5,000</i>	<i>Incomes of \$5,000 and over</i>
1926.....	2.94	12.9
1929.....	3.4	-1.2	19.4
1932.....	-3.8	-2.7	-13.1
1935.....	.71	4.5

* Taken from Table 15, p. 48, of TNEC Monograph No. 4.

Income sources by size of income

It would be very interesting to know in some detail what the sources of incomes of persons in various income groups are, to compare those receiving incomes of \$1,000 with those receiving approximately \$2,000, or \$3,000, and so on. But for income receivers under \$5,000 data are not available in subclassifications. A broad and striking comparison is,

560 THE SOURCES OF PERSONAL MONEY INCOMES

however, possible between those receiving under \$5,000 and those receiving \$5,000 or over. These data are presented as percentages in Table 36—3.

TABLE 36—3 *

SOURCES OF INCOME BY INCOME SIZE

<i>Income class</i>	<i>Wages and salaries</i>	<i>Entrepreneurial withdrawals</i>	<i>Income primarily from property</i>	<i>Total income</i>	<i>Number of income recipients</i>
1926 PERCENTAGES					
Under 5,000.....	71.4	17.2	11.4	100.0	45,517,132
Over 5,000.....	28.5	19.2	52.3	100.0	894,868
1929 PERCENTAGES					
Under 5,000.....	73.5	15.7	10.8	100.0	47,522,929
Over 5,000.....	27.0	16.8	56.2	100.0	1,032,071
1932 PERCENTAGES					
Under 5,000.....	73.8	11.4	14.8	100.0	50,146,558
Over 5,000.....	43.0	12.4	44.6	100.0	356,442
1935 PERCENTAGES					
Under 5,000.....	68.7	17.7	13.6	100.0	51,826,885
Over 5,000.....	37.0	16.0	47.0	100.0	500,115

* Data taken from Table 15, p. 48, of TNEC Monograph No. 4.

Under the heading "income primarily from property" are included interest, dividends, rents and royalties, and realized capital gains and losses.¹⁰ The differences are striking, especially so during the prosperous years of the 'twenties. These figures are clear evidence of the extent to which concentrations of personal incomes in the United States reflect concentrations in property ownership, and hence in incomes from property sources of one kind or another. They are important data in the formulating of social policy relating to income distribution and redistribution and to savings and investments.

¹⁰ Also a few minor miscellaneous items.

CHAPTER 37

Income Distribution and Public Policy

THERE are two issues on which there are probably more widespread disputes than on any other phase of public policy: policies designed to affect the distribution of income in society, and policies designed to prevent or reduce general business fluctuations. Most people agree that business depressions are undesirable, although there have been many political disagreements as to what should be done to prevent them. The issues involved in this area of public policy will be examined in Chapters 44 and 45. Concerning income distribution there are violent disagreements over not only what action should be taken, but also what should be the purpose of action. It is this Pandora's box that is now to be opened.

Many ethical and personal judgments enter into opinions as to what would be a "desirable" distribution of personal incomes. Some differences in opinion rest on deep philosophical foundations that ramify through the whole of economic and social life. Some differences are expressions of the self-interest of particular individuals and groups. It is to differences of the first kind that attention will be especially directed in the first two sections of this chapter. The last two sections will summarize briefly some of the ways in which public policy may affect income distribution.

"Economic Contribution" and the Distribution of Income

Several times in this book there has already been occasion to compare the rewards received by owners of a productive agent with the "economic contribution" made by a marginal unit of that agent. In previous discussions "economic contribution" was identified with the sale value of the marginal product of the agent in each of its employments. It is now time to re-examine the criterion of "distribution in accordance with economic contribution" in the light of the preceding analyses of the distribution of personal incomes. Several interesting problems emerge from such an examination.

What in fact are the economic contributions of labor? Are property incomes accruing to individuals to be regarded as returns for economic contributions of these individuals or are they to be regarded as the income of a society as a whole, derived from the collective resources of the

society? Are interest payments returns for the "sacrifice" of "waiting" by the individual savers providing investment funds? Are pure profits to be regarded as payments for the "economic contribution" of taking the chances in a dynamic economy characterized by many uncertainties? What about "profits of exclusive position"? What of the contrasts between high wages in union-restricted occupations and low wages in many free labor markets? And what of the initial individual inequalities of opportunities in their effects on the economic contributions of different individuals? Finally, does this criterion harmonize with the idea that resources should be allocated in accord with consumer preferences; does the goal of income distribution in accordance with economic contribution provide the income incentives consistent with such an allocation of resources?

"Economic contributions" and purely competitive general equilibrium

The simplest type of situation imaginable in an economy is that presented in the abstract model of purely competitive general equilibrium. It is therefore convenient to begin this re-examination of the criterion of distribution in accordance with "economic contribution" by considering such a hypothetical situation. In purely competitive general equilibrium personal net incomes would be of only two kinds: wages of labor, and interest payments for the use of loanable funds. Two basic questions then present themselves.

1. First is the question of how one might view the discounting of the sale values of the marginal products of labor in a purely competitive economy. Wages tend, under conditions of pure competition, to equal the sale values of the marginal products of workers of each type *minus* a discount of interest payments for the use of loanable funds invested in these labor services between the time at which payments for labor are made and the time at which the resulting products are sold. The workers themselves receive less than the sale values of the marginal products resulting from their employment. There are two principal kinds of ethical evaluation that might be made in viewing these discounts: (a) The interest payment may be regarded as a reward for postponement of consumption on the part of the individual providing the investment funds—in classical economic doctrine this was termed a "reward for waiting." Those taking this view would redefine the "economic contribution" of the labor, considering it to be the *discounted* sale values of marginal products. The interest payment is then the "economic contribution" of the marginal \$100 of investment funds provided by individuals who have the alternative of spending these funds on current consumption. It should be pointed out here that the decision of an individual to use funds for consumption or for investment arises every time a question of replacement of depreciated capital appears. From the point of view of the individual, he is "waiting" to consume later even when there is channeled

back into investment funds just enough for replacement and no more. When there is neither net saving nor dis-saving it simply means that on the average individual decisions between consuming and investing, hence waiting, are such as just to keep the existing stock of productive resources intact.¹ (b) On the other hand, there are some people who would say that this attitude that interest is a just reward for "waiting" is unrealistic and is an ethically false position, that the entire sale value of the marginal product of each type of labor is labor's "economic contribution." According to this view investors are cutting in on what should "rightly" be regarded as labor's "just" reward.

2. Second, there arises an ethical question as to who "should" have claim to any net returns (over and above replacement allowances) that may be included in the sale values of the marginal products of material agents of production. If interest payments are regarded as a "just" reward for "waiting," the answer is simple. These net returns to material agents are then "rightly" claimed by those who provide the loanable funds that are invested. But if this position is not accepted, then again there may be alternative interpretations. This net return may be regarded as fairly claimed by all of the people collectively, to be divided according to some criterion entirely apart from that of the economic contributions of individuals, since the contributions themselves are in this view collective. It may be argued, for example, that such income should be divided evenly on a per capita basis, or on the basis of some other criterion of "need," or possibly in proportion to the labor contribution made by the various individual members of the society.

"Economic contribution" and "uncertainty"

In the analysis of general equilibrium in a purely competitive economy there was no place for the consideration of the role of uncertainty in economic life, and hence of the ethics of income distribution as affected by the changefulness of a dynamic economy. "Pure profits" arise out of such uncertainty. Positive pure profits may be regarded as rewards compensating for the negative profits (losses) of those who take the chances in investing in economic enterprises. The concept of sale value of marginal product (or discounted sale value of marginal product) is inapplicable as a measure of the "economic contributions" of those in a position to receive pure profits.

Again, however, there are two principal ethical points of view. One starts from a basic philosophy of a free enterprise system and the sacredness of private property; from this basic position arises the attitude that profits, positive and negative alike, are "rightly" returns in good or bad fortunes to individuals taking the chances in business enterprise; this chance-taking is viewed as a kind of "economic contribution." The socialist position, on the other hand, settles these positive or negative

¹ It is further assumed that (1) decisions of businesses to save are merely reflections of decisions of individual investors in these businesses; and (2) there are no uncertainties that necessitate that investors take chances.

profits on the whole of the people, who are regarded as the collective owners of the wealth of the society and hence the collective gainers when uncertainties materialize profitably and the collective losers when fortunes are bad.

"Economic contribution" and economic opportunity

The belief that income "should" be distributed according to "economic contribution," however that criterion may be applied and however crude it may seem, is usually closely associated with a belief in the importance of equality of opportunity for the individuals in a society to enter into whatever occupations and whatever business ventures they may individually prefer. To most people it also implies an equality of opportunity to obtain training and acquire knowledge of available outlets for labor services and investment funds. This imagined equality is to an important extent an illusion even in the United States. The accident of birth, whether into a privileged and gifted family or one that is underprivileged and poorly equipped, whether into an area and race and social class where educational and job opportunities are great or where they are small, starts individuals off on very different footings. Monopolistic and monopsonistic situations in both labor and investment markets further block the freedom of opportunity of individuals to maximize their economic contributions. These practices maintain at high levels the contributions made by a few through excluding others from joining them; and they lower the contributions made by those who are excluded since excluded people must find outlets for their labor and investments in relatively oversupplied fields. These problems have already been discussed in some detail in earlier chapters. It is therefore unnecessary to elaborate them further here.

The criterion of economic contribution and the allocation of resources

The criterion of economic contribution as a desirable basis for the distribution of personal incomes is frequently advocated by those who regard this kind of income distribution as the most expedient way of getting resources efficiently allocated and maximizing the national real income. Income distribution is considered primarily as a tool by which incentives in production are most simply provided. Distribution of income according to economic contribution is regarded as "just" in that people are paid according to what they add to consumers' economic welfare through the labor they offer or the material resources in which they invest. Such a basis of income distribution is also regarded as expedient in providing an automatic elimination of the necessity of governmental decisions that involve direct discrimination between different individuals and different groups in the dividing of incomes. It is also consistent with the emphasis on individual freedom that has been so important in the United States; such individual freedom would be difficult to attain (though not inconceivable) within an economic struc-

ture in which the basic determination of income distribution departed radically from the criterion of economic contribution. These are the basic arguments on which proponents of a private enterprise economy rest their opposition to proponents of socialistic schemes.

Need and Equality

Comparison of an important socialist ideology with attitudes in the United States

"From each according to his ability; to each according to his need"—these are goals set by many believers in socialism. Their realization would involve the denial of any right to property income as such, though the society collectively would of course receive such income, which would be divided in some way among its members. It would also involve denial of the right of individuals to receive labor incomes based on their "economic contributions," however defined. Proponents of this position usually assume that "needs" are essentially based on external conditions, as size of family, health problems, and so on, and that all men are equal in the more "psychological" aspects of their consumption needs. These two goals are therefore also very close to a statement that incomes should be divided equally, incentives to productive action taking forms other than income rewards. The fact that full attainment of such goals is impractical, that even in "communist" Russia they were far from realized, is important in the understanding of economic adjustments. But they are nevertheless important as goals for some socialist reformers; and peoples who accept these goals may move in a direction quite different from those subscribing to other hopes and creeds.

The extreme view embodied in this socialist slogan provides an interesting point of departure in considering criteria of income distribution in a nation like the United States. Traditionally in this country great emphasis has been placed on the values of a free enterprise economy and the fundamental equity it is assumed to achieve. In the realm of income distribution this assumed equity has consisted of an approximation to a somewhat fuzzily defined goal of distribution according to "economic contribution," regarding *pure* profits and the interest payments to those providing loanable funds as falling within this equity concept. Failures to attain such a distribution of income are recognized to exist, but are commonly regarded as disfigurements on an economy fundamentally just and sound. Consistent with this position is an objection to such inequalities as may appear because of interferences with the free opportunity of individuals to find outlets where their economic contributions (from both labor services and investments) may be maximized.

But despite this fundamental adherence to the criterion of distribution according to "economic contribution" there has always been recognition of other criteria as well. Those incapable of producing enough to sustain themselves and their families at a minimum of decency in living have been given assistance; and extremes of inequality in the distribution of

incomes have been regarded by many people as undesirable even when these incomes were consistent with the economic contributions of those receiving them. While "need" in the sense implied by the socialists' slogan is very close to a concept of equality, in the United States these two aspects of distribution ("need" and "equality") have been regarded quite separately. By "need" is meant a minimum for all. It is generally accepted that such a minimum is desirable, though this minimum has been only vaguely defined. At the same time, "equality" has generally been regarded as undesirable; such a goal is fundamentally inconsistent with the American focus on the advantages of aggressive private initiative based on an income incentive. The idea that extreme inequalities should be modified is almost universally accepted insofar as these inequalities arise out of positions of exclusion; where they arise for other reasons, opinions are probably quite sharply divided. Some of the implications of these attitudes toward public policy in relation to income distribution in the United States will become apparent in the following pages.

Arguments commonly presented for and against equalization of income distribution

Arguments for a policy designed to bring about a distribution of income very close to equality appear in a variety of forms. By some people it is regarded as self-evident that such a distribution of income is more equitable than any other, no further comment seeming to them to be necessary. Others expand their positions, suggesting that equal distribution of income would remove undesirable pressures to "keep up with the Joneses," that it would eliminate much "wasteful" consumption and shift resources into uses that would be conducive to the building up of a more healthy and efficient population, that equal opportunities can only be attained if children are reared in families having approximately equal income positions, that such equalization of incomes would remove much of the dread and devastation wrought by the insecurity of the individual under existing arrangements. The list of arguments might be extended and elaborated at great length. Most of these arguments are noneconomic; there is little that economic analysis can say that would illuminate further the issues involved.

Arguments against equality in income distribution are also varied and extensive. They include, for example, appeals to the "equity" of higher incomes for those who work hard and contribute much to the society than for those who loaf along and contribute little or nothing. Two of the arguments presented against a high degree of equality in income distribution are especially important in relation to the operation of an economic system. These are (1) that some income inequality is a necessary incentive to get goods produced efficiently, and (2) that unless some people have very high incomes there will be insufficient saving and a society will tend to use up resources without accumulation and possibly even without replacement.

1. Experiments in Russia seem to prove the truth of the first of these arguments against income equalization. Despite their effort toward equalization of income through a centrally controlled economy, the Russians have been forced to introduce income differentials in order to persuade people to take positions of responsibility and to acquire skills. How great these differentials would need to be to get efficient productive activity remains an open question, though it is clear that they need not be as great as the inequalities currently existing in the United States. Moreover, other incentives may to a considerable extent be substituted for income incentives.

2. The argument that if income distribution were very even savings would be seriously curtailed is also of some importance. Savings of individuals provide some of the replacement funds for the maintenance of capital resources intact; they also provide funds for investment that may lead to greater and greater accumulation of capital resources as productive activity is shifted from production for immediate consumption to production of resources that yield consumption goods only through indirect processes and in a more remote future. On the other hand, individual savings are not the only way in which funds may be allocated to investment uses. In the private enterprise economy of modern United States a large part of saving is business saving, earnings of corporations that are reinvested instead of being paid out in dividends. Whether these business savings would be cut off by an equalizing of personal incomes would depend on the methods used to bring about the close approximation to equality. Finally, this problem of curtailed saving is no argument at all if the equalization of income distribution is brought about under a centrally controlled economy, since the allocation of resources between production for immediate consumption and the production and accumulation of capital resources is a matter that can be decided by the dictators in power. Here the Russian experiment with socialism seems to prove the feasibility of at least some accumulating of capital without reliance on individual savings provided controls over these processes are in the hands of the dictator.

Public Policy and Income Redistribution

There are many aspects of public policy that lead directly to redistribution of the real income in an economy; some of these programs are directly intended to accomplish these results; others are initiated for other purposes but indirectly have important redistributive effects. This discussion of income redistribution by public policy is distinguished from the following discussion of public policy and changing income distribution at the source. Here we are concerned only with those policies that take incomes from individuals who have received them and then use the money thus received in such a way as to divide the benefits differently from the collections. Any given set of public expenditures may be financed in any of a number of ways; different methods of financing will

draw incomes from individuals in different proportions, thus having different kinds of redistributational effects. Similarly, any given public revenues may be spent in various ways that will have different effects in the distribution of the benefits received among the members of the society. These matters will be examined in greater detail in Part X of this book, "The Public Economy"; present discussion will therefore be very brief.

The most direct and quantitatively important effects of public policies which redistribute income are of two kinds: (1) They may primarily aid those who would otherwise be below a "decent" minimum plane of living. (2) They may modify or exaggerate inequalities of income throughout the income scale. Public policy which increases or diminishes incomes received from the exploitation of exclusive positions is primarily policy affecting income at the source; though in some cases policies that redistribute incomes are important in soaking up profits of exclusive position. They may also be directed more heavily against property income than income received as rewards for labor services. Since the most important effects of most policies leading to income redistribution are in providing for those at the bottom of the income scale and in modifying or exaggerating inequality, we shall begin by a brief consideration of a way of measuring income inequality. We shall then examine the effects of some different ways of collecting public revenues and some different ways of spending these revenues.

The measurement of inequality of income

Before proceeding to a few comments concerning public policy and income redistribution, it is well to lay a background for discussion by

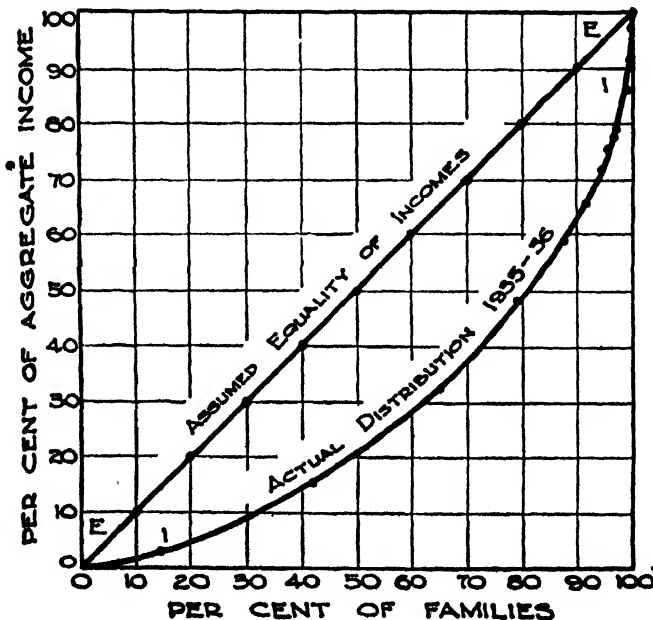


Fig. 37—1. Measure of income inequality.

introducing a measure of the degree of inequality of income. Such a measure is provided in a very simple graphic technique.

Suppose the \$48,000,000,000 of income received by families of two or more in the year 1935-36 had been divided evenly among the 29,000,000 families of the country. This would have meant an income of \$1,655 per family. Ten per cent of the families would receive 10 per cent of the total national income, 20 per cent of the families would receive 20 per cent of the total national income, 30 per cent of the families would receive 30 per cent of the total national income, and so on. This absolutely equal distribution of income by family units (though not per capita) could be represented graphically by constructing a chart; along the vertical axis would be marked off per cents of the national income,

TABLE 37—1

PERCENTAGE DISTRIBUTION OF FAMILIES AND AGGREGATE INCOME RECEIVED

<i>Income Level</i>	<i>Per cent of families</i>	<i>Per cent of income going to these families</i>
Under \$500	14.21	3.01
Under 1,000	41.68	15.57
Under 1,500	64.63	32.89
Under 2,000	79.05	48.08
Under 2,500	87.43	59.56
Under 3,000	91.90	67.04
Under 3,500	94.43	72.04
Under 4,000	95.92	75.45
Under 4,500	96.77	77.65
Under 5,000	97.29	79.16
Under 10,000	99.03	86.52
Under 20,000	99.68	91.79
Under 100,000	99.99	98.13
All levels	100.00	100.00

along the horizontal axis per cents of the families. A straight line drawn diagonally from the lower left hand corner to the upper right hand corner will then indicate the per cent of families receiving the associated per cent of the national income. Such a situation is illustrated in Figure 37—1.

Now let us compare this with the distribution that actually existed in 1935-36, using figures for all families, including incomes received as relief payments. The data are given in Table 37—1: 14.21 per cent of the families had incomes under \$500 and received 3.01 per cent of the national aggregate of family incomes, 41.68 per cent had incomes under \$1,000 and received 15.57 per cent of the aggregate, and so on. These data are also plotted in Figure 37—1, but instead of making a straight line they make a curve, which lies close to the base line at first and then climbs steeply at the right. The depth of the depression in this curve is a rough index of the degree of inequality in this distribution of income.

Redistribution and the collection of public revenues

Public revenues may be obtained by issuing new money, by borrowing, and by taxation. Some of the ways in which revenue collections redistribute incomes are easily illustrated by considering just four important taxes: the personal income tax, the general sales tax, the taxes on corporate profits, and the property tax. These taxes will be considered only briefly here; they will be examined more fully in Chapters 48 and 49.

The personal income tax as it is set up by the federal government in the United States today takes a larger per cent of income from those whose incomes are large than from those whose incomes are small. It may thus be used as a method for redistributing income in favor of the lower income groups, diminishing inequalities in incomes as received through private enterprise. The general sales tax, on the other hand, takes a larger proportion of small incomes than of large. It falls only on that part of income that is spent on consumption goods, and since those with low incomes spend most of what they have while those with large incomes save a lot, the "little fellow" pays a general sales tax on a much larger per cent of his income than does the wealthy man. The general sales tax, therefore, has the effect of increasing the degree of inequality of personal incomes. Redistribution of income is part of the intent of many of those who favor personal income taxes, especially income taxes that take much larger proportions of income from those with big incomes. Redistribution of income is never argued as a reason for passing a general sales tax, and such redistribution is probably not the usual intent of those urging such taxes, but it is the result none the less. Thus there are in the United States important taxes intended in part to modify inequalities of income distribution and other important taxes that however intended in fact exaggerate those inequalities.

The effects of a steeply graduated income tax or of a general sales tax on the distribution of income may be illustrated graphically. Assuming at the start an income distribution such as existed in 1935-36, a steeply graduated income tax falling heavily on high income groups might result in a decrease in degree of inequality such as would be represented by the second line in Figure 37-2. On the other hand, a general sales tax might have the results indicated by the fourth line in Figure 37-2.

The fact that back of much of our thinking about income distribution is the criterion of distribution according to "economic contribution" is illustrated in arguments for taxes designed to capture monopoly profits. Though prevention of such gains in the first place is generally regarded as a preferable procedure, where prevention fails, monopoly profits may sometimes be partially recaptured by taxes. Heavy taxes on incomes of corporations are frequently supported on these grounds, since in many cases they soak up monopoly profits before these profits can be passed out to stockholders as dividends. Such taxes fall on all stockholders alike for each unit of stock held, and regardless of the total income of the individual holding the stock.

The property tax is the chief source of local revenues in the United States. By taxing such property according to its value, the government recaptures some of the property income produced each year. This income is thus shifted from investors to the general population. Taxation of this sort modifies income inequalities insofar as those receiving incomes from the taxed property are in the higher income brackets; while this is generally the case, it is not always so. Those who argue that the productive contributions of property are essentially collective rather than

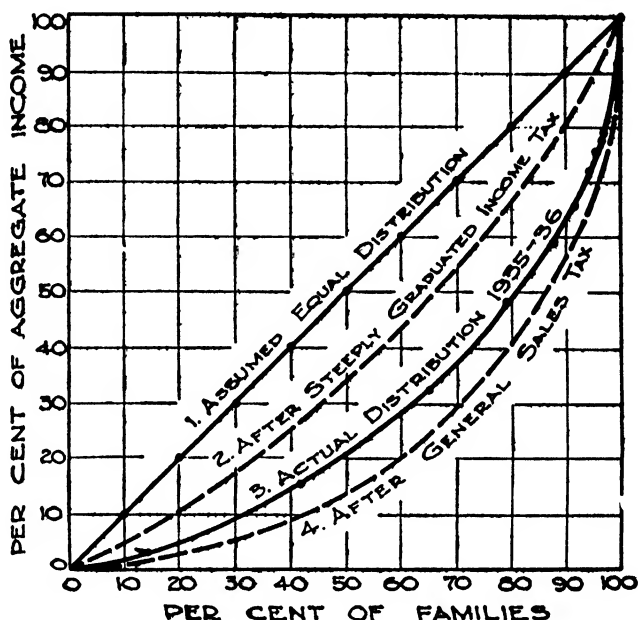


Fig. 37—2. Effects of selected taxes on degree of inequality of incomes.

the contributions of the individuals owning the property defend such taxation as consistent with an income distribution according to individual economic contributions; others hold the opposite view.

Redistribution and public expenditures

All public expenditures lead to some redistribution of incomes, but the most important effects may be analyzed by considering three principal groups of expenditures: (1) payments of relief money (or grocery orders, and so on) to families in very poor circumstances, and provision of services such as public health clinics that are made available only to low income groups, (2) provision of services such as police, highways, and schools, on an equal basis to all members of the population, and (3) payments of subsidies to special producer groups, as illustrated in the payments to farmers in the agricultural program.

While most of these public expenditures tend to equalize incomes some-

what, this is not always the case. A few simplified illustrations will help clarify these situations.

1. Most obvious is the equalizing effect of payments in relief money and of services made available to low income groups only. Such policies increase the incomes of low income groups and hence their per cent of the total national income. They do not exert in themselves any significant effect on inequalities as reflected in concentrations of incomes at the top. The effects of such policies as applied in 1935-36 are already included in the data of Table 37-1.

2. Services provided on equal terms to all members of the population will have a more or less important effect on income distribution depending on the proportion of total real income in an economy that is received through such government services instead of through production in private enterprise. If such services are of great importance they may flatten out the distribution markedly despite remaining inequalities. Some services ostensibly provided on equal terms are in fact used quite unequally by different income groups. The highest income groups may fail to make full use of public school facilities, for example, thus increasing the equalizing effects of these services. On the other hand, some services, such as the national park recreation areas, are used primarily by those in the middle and higher income groups and are inaccessible to those on very low incomes; such services therefore increase inequality between the lowest and the middle and upper groups.

3. When subsidies are paid to special groups of producers effects in modifying or exaggerating inequality will depend on two things: (a) whether these payments go in larger proportion to the low or the high income families *within* the group aided, and (b) whether the group as a whole tends to come in the lower or upper parts of the income scale. Benefits under the agricultural adjustment program in the 'thirties went in larger proportion to high than to low income farmers, and hence tended to increase inequalities within the farming group itself; but they may not have exaggerated inequalities in the total population, when income receivers in all lines of activity are taken into account.

Controlling Income Distribution at the Source

Income distribution may be controlled at the source through policies that affect the sizes of the incomes people receive in the first place. Policies of this sort inevitably crop up in any kind of economy whether characterized primarily by private enterprise, or by central planning. Such policies are very different in character, however, in an economy characterized predominantly by free enterprise as compared with an economy which is predominantly socialistic. In a private enterprise economy many such policies are closely coordinated with concepts of equity in income distribution based on the "economic contribution" of the income receivers; while in a socialist economy the focus is more commonly on an approximation to equality.

Control within a private enterprise economy

There are five major groups of policies that could be especially important in controlling income distribution at the source in the economy of the United States: (1) policies directed against restrictive practices of business, (2) policies directed against restrictive practices of labor, (3) wage legislation (and in wartime, price controls), (4) legislation in the interest of particular industry groups, and (5) policies intended to check general business fluctuations and to lift an economy out of a state of general depression. The last of these will be deferred for the present, since they will be taken up in later chapters; such policies if successful affect income distribution primarily by stabilizing distribution somewhat over time and eliminating some of the insecurities of individuals and groups.²

1. The federal antitrust program is directed primarily against restrictive practices by business enterprise. This program is focused on a freeing of the channels of trade in order to increase production and thus raise the national real income; but insofar as such a policy is enforced there are several distributional effects associated with the increased production it brings about: (a) It tries to prevent the occurrence of "profits of exclusive position." (b) It attempts to increase income opportunities for many laborers and investors by permitting them to enter an industry that in the absence of antitrust action would have been closed to them. (c) By leading to expanded outputs and thus to lowered prices in the industries against which proceedings are brought, antitrust action increases the real incomes of consumers as a whole, and hence of the general population. Antitrust action is a major cog in a program designed to bring about an approximation to the goal of distribution according to economic contribution. Since many of the existing inequalities in the distribution of personal incomes reflect monopolistic business practices, a vigorously enforced antitrust program would tend to bring individual incomes somewhat closer together.

2. No clear public policy toward restrictive practices of labor groups has yet taken form in the United States, though the problem has occasionally been attacked on its fringes.³ There are many people today who regard this as a major gap in the antitrust program, and considerable thought is being devoted to methods for handling the problem. A public policy that resulted in the breaking down of restrictive labor practices would, like antitrust policy in business, both increase the national real income and alter income distribution. It would increase opportunities for workers previously excluded from employments where wages and economic contributions were high, while the competition of these newcomers

² In addition to these five groups of policies, any program that redistributes current income or property holdings (especially taxes on inheritances) will affect the distribution among individuals of claims on property income in the future.

³ See Chapter 33.

would lower the wages received by workers in the previously privileged positions. The result would be some modifying of inequalities in wage incomes, and an increase in the economic contributions of the previously excluded workers.

3. The effects of wage and hour legislation were discussed in Chapter 34. If legally established minimum wages are set very high, the result in competitive labor markets is likely to be some unemployment, raising some individuals previously in extremely low income groups up toward the middle, but pushing others onto relief. Such a program would therefore only be feasible if accompanied by a supporting public relief program, whether direct relief or work relief. The redistributive effects of the relief program are thus mixed up with minimum wage legislation. When minimum wage legislation leaves some jobs uncovered, it may push more workers into uncovered fields and further depress wages there. On the other hand, when minimum wages are enforced in previously monopsonistic labor markets the result may be even increased employment, the curtailing of the "profits of exclusive position" of the monopsonistic firms, and a diverting of these profit incomes to wages. This would bring workers' wages more in line with their economic contributions. Thus where wage legislation destroys monopsonistic situations it will probably tend both toward some modifying of inequalities of personal incomes, and toward a closer approximation to distribution in accord with economic contribution.

4. Special privileges to particular industry groups take many forms.⁴ There may be a tax imposed on oleomargarine in the interests of butter producers, tariffs on sugar from Cuba in the interests of domestic beet-sugar producers, or any number of similar policies.

Where there is freedom of entry into an industry, special privileges granted to its members have different effects in the short and in the long run. In the short run they lead to a shift in income distribution in favor of investors in the industry and against the rest of the population. In the long run new firms will appear in the privileged industry and returns to its members will fall back to the level that they would have reached without the public action. The principal long-run effects are therefore a lower national real income and a changed use of resources. The implications of this analysis will be developed much more fully in later discussions of tariffs.

Where there are restrictions on entry into an industry, special privileges granted its members will have more permanent effects on income distribution. They will raise the incomes of investors in the privileged industry, exaggerating the profits of exclusive position. This will take place at the expense of the rest of the population, who will suffer a reduction of real income greater than the gain of the privileged monopolists. Since those

⁴Subsidies are a part of such special privileges. They involve a direct redistribution of income, and were therefore discussed in the last section. It is only as they are part of a broader plan that affects incomes received in the sales of products or services that they play any role in altering income distribution at the source.

receiving profits of exclusive position are typically in the high income groups, a policy that increases their incomes at the expense of the rest of the population is a policy that will exaggerate inequalities in income distribution.

When special privileges to particular producer groups become very widespread they may approximately cancel each other out in their distributional effects, though it is more than likely that they will exert a dampening influence on the total level of economic activity and the national real income.

Income distribution in a socialistically organized economy

In any economic system, whatever its structure, decisions must be made as to what shall be produced, something must be done to get these things produced, and there must be some technique for dividing the resulting products among the individual members of the society. These fundamental facts were stated in Chapter 2 and have been illustrated throughout the pages of this book.

In a private enterprise economy the incentive to get goods produced is primarily an income incentive; and the allocation of resources and the distribution of incomes are in large measure determined simultaneously through the same pricing adjustments in the markets. In a socialist state these two aspects of economic adjustment are much less closely tied together. The central planning board of a socialist state may dictate production plans including what shall be produced and by whom and how incomes shall be divided, the board's dictates displacing the price mechanism of the market; or the planning board may permit many adjustments to be made through a market mechanism approximating that of a private enterprise economy, but rigorously control the operation of that mechanism. The board might, for example, rely to a considerable extent on market pricing as a guide to what to produce, but pay out incomes to individuals without any regard to their economic contributions. In the most extreme case it is even conceivable that the central planning board might make no money income payments at all; individual incomes could then take the form of ration cards giving people the right to certain quantities of each specific commodity. There is no single or simple blueprint of what a socialist state is or "should" be.

While in a centrally controlled economy the problems of resource allocation and of income distribution are more readily separated than in a private enterprise economy, even the dictator cannot completely ignore income incentives in getting goods produced. Neither can he ignore completely the consumer preferences of the many individuals who make up the population. In any practical situation a predominantly centrally controlled economy will inevitably have elements resembling private enterprise, just as a predominantly private enterprise economy will have elements of central planning.

How will the policies of a socialistically controlled economy relate to the criteria of income distribution discussed in the first sections of this

chapter? This question can be answered only in rough terms, since there are different points of view that may be developed in different socialistic economies. Two general sets of ideas do, however, have wide application: (1) Insofar as rewards in accord with economic contribution are regarded as "equitable" the interpretation of economic contribution is labor contribution only. Property incomes are regarded as the contribution of the collective wealth accruing to the people collectively. If such wealth is owned collectively, income incentives are not needed in order to insure its use in the most productive or, from the point of view of the central planners, "desirable" ways. Income incentives are therefore confined to incentives to labor. (2) Great emphasis is placed on approximating equality in income distribution. Planners in a socialist state will therefore tend to gear their income distribution policy to this goal insofar as is practicable.

Part VIII

MONEY, CREDIT, AND PRICES

PREAMBLE

FOR many decades economists regarded money as a subject of relatively minor importance, to be treated for the most part quite independently of the major body of economics. Analysis was often conducted on a "real" plane in order to draw aside the "veil of money," which might obscure the real problems of production and distribution of wealth and income. In recent years many economists have swung to the other extreme, asserting that money and monetary changes constitute the core of all economics and that it is only by properly controlling money that we can achieve a smooth-working and effective system of production and distribution. Many of the "real" problems that occupied the older economists are relegated to positions of subsidiary importance by these newer writers.

As is so often the case in the historical development of ideas, both groups have much to contribute to an understanding of the problems at hand and both have overstated their respective positions. It is now clear that many of the so-called "classical" economists seriously underestimated the importance of monetary instability in affecting the general level of employment and production. On the other hand, it is also clear that many modern writers have gone much too far in supposing that we can cure the evils of general unemployment alone by tinkering with the flow of money and credit. We should never let ourselves be blinded to the ultimate importance of "real" as against "money" income. Yet the developments of recent years have demonstrated beyond reasonable doubt that only by controlling the flow of money income better than we have done in the past can we expect to maintain conditions of prosperity and reasonably full employment. It is toward the discussion of general unemployment, inflationary booms, and their remedies that the study of money in this Part is largely directed.

CHAPTER 38

Introduction to Money and Credit

Money in a Complex Society

THE word "money" ordinarily connotes a miscellaneous assortment of coins and bills, of which there hardly ever seem to be enough. We know that money enters into almost every aspect of life—we are paid our incomes in money and we pay out this money for most of the goods and services that we purchase. The activities of the business world are based to a large extent on contracts expressed in terms of money; and the fundamental purpose of most businesses is to obtain a money profit by obtaining total money receipts in excess of total money expenditures.

To most of us getting money seems to be the primary objective of economic activity—the "mainspring of economic motivation." Actually, of course, most of us want money not for itself but for the command it gives over goods and services. It is obvious that in such a complex economic system as ours the use of money is of fundamental importance in avoiding the awkwardness of a barter system. But of perhaps equally great importance and much less obvious is the central part money plays in the drastic fluctuations in business activity, employment, and prices that we know as booms and depressions.

The major functions of money are ordinarily said to be (1) as a medium of exchange, and (2) as a standard of value in terms of which different goods and services may be compared. The usefulness of money in these two respects is obvious. Comparison with a barter economy indicates the tremendous difficulties a moneyless economy would face. Suppose Mr. A has produced a pig and he as a consumer wants a spool of thread, a movie, three pairs of socks, and a weekly newspaper. He may then start out to barter his pig (presumably butchered) for these various objects. He may have heard that B has produced some thread, but unless B wants pig in exchange, A is still out of luck. Perhaps C wants a pig and offers cotton cloth in exchange. If A trades his pig for cotton cloth and can then find someone who will take the cotton cloth in exchange for thread, A's problem is solved, so far as the thread is concerned. But the difficulties of even such a simple transaction by barter indicate the magnitude of the problem of conducting all exchange in that manner.

A partial solution to A's difficulties might be found by establishing a market, where all persons wishing to exchange would meet regularly. But

even if A managed to find others willing to exchange thread, movies, socks and newspapers for pig, there would still be the problem of the exchange ratios between these various objects. When we have things to exchange we ordinarily convert them into money values, and so compare them in terms of a common standard of value. But under a purely barter system even this standard of comparative values would be lacking.

In addition to these two major functions of money, two others are often listed: (3) as a standard of deferred payments, and (4) as a "store" of value. We want not only to compare the values of different things at the same time, but also to compare the values of things now and in the future. Money is useful in providing a standard for deferred (future) payments in this respect. And not only do we often want to compare future with present values, but we also want to store up goods and services for future contingencies. We might simply store physical goods; but this is a troublesome and expensive procedure in many cases. Furthermore, often we want merely to store up purchasing power for the future, rather than any particular good. Money is of great use in providing such a "store" of value or purchasing power, but, as we shall see, this use sometimes leads to all sorts of trouble.

These advantages of the use of money in a complex economy may be stated in a slightly different manner, emphasizing the necessity for the use of money if the aims of efficient use of resources and their allocation in accordance with consumers' preferences are to be realized.

1. *The use of money permits consumers to receive their incomes in the form of generalized purchasing power and so enables them to select freely from among the various goods and services offered for sale.* In a barter economy, even with a very small number of products considered, the channels for expressing consumers' preferences are blocked by the limitations of the specific objects that they have to offer in exchange. In a money economy, everyone is willing to sell for money, because he knows that he can, in turn purchase other goods and services with the money. Therefore, the consumer having money (generalized purchasing power) can express his preferences freely simply by offering money for what he wants.

2. *The use of money permits consumers to spread consumption over time with less regard to the precise dates on which they receive their incomes.* It is much easier to store money than goods.

3. *Money provides a measuring rod by which producers may compare the advantages of producing one good rather than another.* In a barter economy there would be no uniform standard registering consumer preferences for one good over another. Adjustments of production to consumers' demands occur only in terms of a whole range of exchange ratios between concrete goods, and each of these ratios is difficult to discover except as revealed by isolated instances. In a money economy consumers' preferences are expressed in uniform "dollar votes," which serve as guides to producers.

4. *By facilitating exchange between goods and by expediting payment over time, money makes possible the use of productive techniques involving roundabout and highly specialized processes.* In a very simple society where wants are relatively uniform, goods few, and where most production is for home use, a system of direct barter might conceivably suffice. A complex system in which consumers and producers are far removed from each other in time and space is only conceivable with the introduction of money.

These advantages deriving from the use of money in facilitating the smooth functioning of a complex exchange economy are easily seen. Much less obvious is the part money and monetary changes play in the booms and depressions that have characterized modern economic systems. The use of money has been taken for granted all through the discussion of the allocation of resources and distribution of income, and no particular problems arose because of failure to make specific the use and advantages of money. But there was throughout a tacit assumption of the absence of booms and depressions—the absence of widespread unemployment of men and machines. It is to this problem of business fluctuations that the discussion of money and credit is primarily directed, since monetary changes play a central part in explaining the cumulative expansions and contractions which have been so disruptive to the smooth working of modern economic and social systems.

"Money" and "Credit"

We are used to thinking of money in two forms: circular pieces of metal (coins) of particular size and design; and rectangular sheets of green paper (bills), identified by particular pictures and signatures, plus a number to tell the denomination of the bill. But for hundreds of years cattle served as money in the ancient world. In the late Roman empire small square pieces of leather were used as money. As recently as two hundred years ago hides and skins, wampum beads, cowrie shells, and numerous other such objects served as money in the North American colonies. It is only recently that the world has widely adopted the types of coins and paper notes with which we are familiar; and today there is strong evidence that we are moving more and more away from the use of metal coins for any except very minor purposes. It is erroneous therefore to suppose that there is something inherent in the things that we call money that necessarily makes them money. By the same token, it is erroneous to suppose that money can be defined merely in terms of the substances that we happen to use for it.

Any useful definition of money must run in terms of what money does, not what it looks like. A fairly common definition may be adopted: money is anything that passes freely from hand to hand in an economic or political area as a means of payment of obligations irrespective of the credit of the person presenting it. Thus the dollar bill with which one

pays for a pair of stockings is, of course, money. But if you were to pay for the stockings with a personal check upon your bank account, this check would not be money under the above definition because the sales-girl would not accept it unless she had reasonable assurance that your credit was good (that is, that the bank would pay your check). This check we should then classify as a "credit instrument." Credit instruments are essentially identical with money in function except that they are not so generally acceptable as means of payment. In modern times Americans have become so accustomed to using bank checks as money that for all practical purposes payment by bank check is the equivalent of payment by actual money. In fact, payment by bank checks is quantitatively much more important than payment by actual money in this country today. Although a formal distinction between "money" and "credit" is convenient for discussion, it is important to recognize that this distinction is largely formal—that money and credit are essentially the same so far as making payments are concerned. Indeed, in some cases (for example, a "cashier's check") it is very difficult to draw the line between the two. When the various types of money and of credit are discussed briefly in the following sections, it will become apparent that the separation is fairly easy for the types most commonly used; but later the essential sameness of the functions fulfilled by money and credit will be emphasized. It is useful, therefore, both to distinguish them and to see their likeness.

Types of Money

Table 38—1 indicates the major kinds of money in use in the United States around the middle of 1942. Part of this money is at any time in

TABLE 38—1

TYPES OF MONEY IN THE UNITED STATES ON JUNE 30, 1942 *

<i>Type of Money</i>	<i>Amount</i>
Federal Reserve Notes	\$ 9,310,000,000
Silver certificates	1,754,000,000
Silver dollars	66,000,000
Subsidiary silver	504,000,000
Minor coins	213,000,000
United States Notes (Greenbacks)	317,000,000
Others	218,000,000
Total	<u>\$12,383,000,000</u>

* Data are from *Federal Reserve Bulletin*, September 1942. All figures are rounded to the nearest million. "Others" is composed of national bank notes and other forms of currency now being retired from circulation.

the vaults of the banks and part in the hands of the public. Silver certificates are \$1 and \$5 bills, which represent that much silver stored in government vaults. They are essentially warehouse receipts for silver

which is held by the government. Subsidiary silver consists of half dollars, quarters, and dimes. Minor coins are nickels and pennies. Federal Reserve Notes are paper money, varying in denomination from \$10 to \$10,000, issued by the Federal Reserve Banks. Their nature is discussed in a later chapter. Here it is sufficient to note that the law requires that at least 40 cents in gold be kept as partial backing for each dollar of Federal Reserve Notes. It is apparent that the actual metal in circulation is a very small fraction of the total money supply, and only a small part of paper money is required to be "backed" by gold.¹

Until 1933 small numbers of gold coins circulated, primarily for use as gifts, lucky pieces, and so on; but in 1933 as a part of the emergency legislation all monetary gold was purchased by the federal government (that is, people were required to turn in their gold coins for other money), and there is little likelihood that gold coins will return to circulation. Moreover, although the vast amount of gold owned by the federal government serves in a general way as a security for the money outstanding, there is little direct relation between them, and gold cannot now be obtained by individuals in exchange for other money except for use in international transactions or in the arts. Except in international trade, gold no longer serves directly as money, and its connection with circulating money, though existing, is very indirect. The main function of gold in this respect has become one of maintaining confidence in the outstanding money issued by the government, and even in this function it is less important than it used to be.²

The fact that the possession of money gives command over goods and services sometimes leads to a confusion between money and wealth. Suppose Mr. X has both a dollar bill and a dollar sack of flour. Suppose now that the sack of flour is destroyed (let us say in a fire)—his wealth has been decreased by the loss of the flour and society's wealth has been similarly decreased. But suppose instead his dollar bill is burned up. His individual wealth would again have been decreased by one dollar, but the wealth of society (its real goods and services) would be just as great after as before loss of the dollar bill. So although we ordinarily speak of an individual's wealth as so many dollars, it is not very useful to speak of society's wealth in this way. In the last analysis, the wealth of society is the goods and services it possesses, and the creation or destruction of paper money does not affect the real wealth of society

¹ To say, as is sometimes done, that any type of money is "legal tender" means that the law states that this kind of money must be accepted *at face value* in payment of obligations. It does not mean that the debtor is freed from his obligation if the creditor refuses to accept the legal tender money; it merely means that the debtor cannot be forced to offer any other kind of money and that he is not liable for interest or collection fees on the debt after legal tender has been refused. Until 1933 only some kinds of money were legal tender in this country, but since 1933 all kinds of money have been legal tender, even for obligations contracted before that date.

² For a more detailed discussion of gold and the gold standard, see Chapter 42.

except indirectly through its effect in stimulating or retarding the production of goods and services.³

Types of Credit

Although many people think of money as the primary means of payment for goods and services, actually over 75 per cent of all the monetary payments made in the United States are made by credit instruments, chiefly bank checks, and not by "money" at all.⁴ Thus, something over \$500,000,000,000 in checks were written in a recent year, whereas payments made by actual money were around \$75,000,000,000. Since bank checks make up so large a percentage of the total payments made (that is, are so large a portion of the total circulating medium), it is important to understand the factors determining the total value of checks that people can write; and this is the subject of the next chapter. Since checks and actual money serve the same purpose, an increase in the dollar value of checks that people can write has the same general effect as printing a like number of new dollar bills. The amount of bank checking deposits may be spoken of roughly as the "quantity of credit" at any given time, since the total of the deposits shows the value of checks that people can write and since check payments bulk so large in the total picture.

Of course there are various types of credit and credit instruments. When a man buys a suit, he may pay in actual money, he may write a check, he may "charge" it, or he may give the merchant a written promise to pay in installments or on some future date, to cite only the most important possibilities. Of the three types of credit, check payment is the closest substitute for money. The merchant receiving the check can take it to the bank and receive money for it immediately if he wishes, though actually he is much more likely simply to deposit the check to his account, so that no actual money changes hands. The check is a promise to pay "on demand" at the bank. If a buyer gives the merchant a written promise to pay at some future date, this differs from the check primarily in the time the merchant must wait before getting the funds promised. A "promissory note" is a less close substitute for money than the bank check. If the buyer "charges" the suit, there is no tangible credit instrument to show his promise to pay. The merchant simply records the debt on his books. Such a debt is a still more distant substitute for actual money. But in all cases the transaction involving

³ The increase or destruction of metal money, such as silver dollars, is a somewhat different matter. The silver in the dollar is real wealth, but it was also real wealth before being coined. The mere fact that it has been stamped as a coin does not increase the wealth of society; and if the coin should be destroyed, it would be the loss of the silver that decreased society's wealth. The loss would be the same whether or not the metal had been coined.

⁴ This estimate, and the others following, are rough, but are accurate enough to illustrate the great relative importance of bank credit.

sale of the suit has been consummated, in one by the use of money and in the others by the use of money substitutes (credit).

In the business world there are a great variety of types of credit—promissory notes, bills of exchange, open-book accounts, and so on. Although these are of importance and differ somewhat among themselves, for elementary purposes attention may safely be centered on bank credit and the other types of credit largely neglected. By so limiting discussion some inaccuracies will arise, but bank credit is so much more important than all others and the conditions determining its quantity so different that more is gained than lost by dealing primarily with it alone.

CHAPTER 39

The Banks and the Quantity of Credit

CREDIT is by far the most important means of payment in the American economy today, and among credit instruments bank checks are predominant. Since bank credit is of such major importance and since the supply of bank credit hinges on the whole process of commercial banking, it is worth while to examine in detail the connection between the banks and the quantity of credit before proceeding to discuss monetary fluctuations and their connection with business fluctuations. Banks act (1) as middlemen between borrowers and lenders, and (2) as "creators" or "producers" of the credit that they make available to borrowers. The first of these functions can be discussed rather briefly; it is the second that needs most attention.

Middlemen Between Savers and Borrowers

Banks as middlemen between savers and borrowers

There are now in this country about 15,000 banks. In 1920 there were above 30,000. The tremendous reduction represents mainly bank failures, which were a steady phenomenon even throughout the prosperous 'twenties and in 1931-33 reached very great proportions. Of the 15,000 existing banks, about 500 are "mutual savings banks"¹ and about 14,500 are "commercial banks," although almost all commercial banks have savings departments. Of the commercial banks, some 5,100 have received their charters to do business from the federal government and are called National Banks. All of the savings banks and over 9,000 of the commercial banks have received their charters from various state governments and are called State Banks. A small number of banks are unincorporated or "private" banks.

One major function of banks is the conversion of savings into investment—that is, making it possible for savings to be invested in productive enterprises. This function constitutes the major activity of savings banks, and is also carried on by commercial banks as well as by other investment institutions. Suppose a saver deposits his savings in a savings

¹ Most of these savings banks are in New England, where thrift and conservatism have been very strong.

bank, receiving a certain rate of interest. The bank then invests these savings, possibly by purchasing government or industrial bonds, or by making loans to farmers to improve their lands and buildings, or in various other ways. Suppose the savings bank uses the funds to purchase a new \$1,000 bond of a public utility corporation. The public utility thus receives \$1,000 of the savings of the original savers, with the savings bank acting as an intermediary. The money savings are converted into physical capital goods when the public utility spends the \$1,000 for machinery or for helping to finance a new building. Receiving interest from the utility on the bond and the \$1,000 principal when due, the savings bank is able to pay the somewhat lower interest rate owed to its depositors and reimburse them in full if required by selling the security. Savings banks merely make deposited savings available to borrowers; they do not increase the amount of credit in existence. Nor are deposits in savings banks means of payment, since checks cannot be written on them.

Savings banks are not the only, or even the most important, agencies for conversion of savings into investment, though they do clearly illustrate the conversion process. The savings departments of commercial banks also receive relatively long-period savings, which "time deposits" the banks then loan out as do savings banks. By contrast, the checking, or "demand," deposits of a commercial bank are thought of as short-period savings, likely to be withdrawn at any moment by the depositor. In keeping with this distinction, the law requires that banks keep larger reserves behind demand deposits than behind time deposits, and permits the bank to require that a saver give advance notice (often thirty or sixty days) before he can withdraw his savings account, although this law is seldom invoked. Savings and time deposits, unlike demand deposits, are not transferable by check and hence do not serve as a medium of payment. But actually the distinction between time and demand deposits is often not very sharp, and it is erroneous to think of savings accounts as only long-period savings moving into investments while demand deposits represent short-term loans. Indeed, as will become evident, demand deposits for the most part are "created" when the banks make loans and investments rather than arising from any saver's depositing his savings at all, and no rigorous distinction exists between demand and savings deposits, either as to origin or as to use.

A somewhat different banking function is performed by "investment bankers." Investment bankers are essentially marketers of corporate securities, rather than bankers in the ordinary sense of the word. Corporations wishing to sell securities ordinarily do not themselves offer the securities for sale. Instead, they engage investment bankers to market the securities. Usually the investment banker or banking syndicate buys the securities outright from the corporation and sells them to the public at a slightly higher price, keeping the difference as payment for marketing services. The investment banker does not receive deposits from the public and invest them—instead he arranges for the sale of stocks

and bonds to banks, insurance companies, individuals, and others who have funds to invest. He is a middleman in the process of converting savings into investment, but his service is somewhat different from that of the regular banker.

"Nonbanking" middlemen between savers and borrowers

In addition to banks, there are numerous other agencies through which savings may be made available to borrowers and converted into investment. One of the most important nonbanking agencies is the insurance company. A considerable part of our total savings is made by buying insurance. Insurance companies receive regular premiums (savings) from policyholders, and invest these funds, usually buying bonds or loaning on real estate. Thus savings are converted into investment very much as through savings banks.

A large volume of savings is converted into investment without the services of any outside agency. If a person buys a stock or bond he is directly converting his savings into investment.² Tremendous issues of stocks and bonds are sold by corporations and governments to individual and institutional investors other than banks. Another important method of direct conversion of savings into investment is the practice followed by most business firms of "plowing back" or reinvesting a part of each year's earnings in expanding plant and other equipment.

Special agencies have grown up to make available savings to special groups of borrowers. Agriculture, for example, has several special types of middlemen of this sort, some of which have been set up by the government. Among the private sources of farm credit, loans by individual persons and farmers' cooperative organizations supplement those of banks and insurance companies. Public credit agencies have grown greatly in importance in recent years. Twelve large Federal Land Banks exist especially to finance the farmer, and in 1933 the Federal Farm Mortgage Corporation was set up by the government to facilitate refunding of farm mortgages at lower interest rates. Since these federal agencies have offered credit to farmers at very low interest rates, they have gone a good way in eliminating private lenders from the field, though insurance companies and private banks still make a large volume of agricultural loans. But even though the federal banks extend the loans to farmers, individual savings are being converted into funds for farmers. The government has generally borrowed the funds and is now loaning them out to farmers. Unless the government has obtained the funds through borrowing new credit "created" by the banks, the process is similar to that of the private savings³ investment process.

Loans made to consumers do not represent conversion of savings into investment goods, because the funds are used for consumption purposes.

² Although the investment banker is likely to have been a middleman in the process.

Yet from the point of view of the saver they are invested. Numerous special agencies for financing consumers also exist. Most important are:

1. Personal finance companies
2. Personal loan departments of commercial banks
3. Insurance companies (making loans to policyholders)
4. Individual "loan sharks"
5. Pawnbrokers
6. Morris Plan Banks
7. Cooperative credit unions
8. Installment sales plans (either directly by sellers or by special installment finance agencies)

The effective rates charged consumers by these various agencies range from 3 per cent or 4 per cent to 42 per cent a year, the upper limit presumably being set by a federal statute. Many of these agencies obtain the funds they loan out by borrowing from commercial banks, so the ultimate savers are the same as in the case of direct bank loans. Like special agricultural credit institutions, consumer credit agencies illustrate the multiplicity of special quasi-banking agencies which have grown up to meet special needs. The list might be extended, but such agencies are not the primary interest of this chapter.

Creation of Credit by Commercial Banks

The distinguishing feature of modern commercial banking is its ability to create and maintain demand deposits (checking accounts), which serve as means of payment. Obviously, savings banks, insurance companies, investment bankers, and such middlemen between savers and borrowers "create" no credit—their activities do not increase the total quantity of means of payment. But the commercial banks (that is, the banks we all know and deal with) operate quite differently, and regularly "create" or "produce" bank credit that serves the same purpose as money. They "loan out more than we have put in them."

Potential creation of credit by an individual bank

In understanding this rather startling statement that commercial banks "create" credit, it is well first to be clear as to exactly what is meant by bank "deposits." The word is misleading for it suggests that deposits originate through storing or depositing actual money in the banks and that the banks always keep in their vaults money equal to the amount of their deposits. Neither of these suggestions is true. It is not true that deposits are simply "money put in the bank." A bank deposit is a *claim* on the bank, giving the depositor the right to demand money from the bank. Demand deposits are simply debts owed by the bank to its "depositors," which debts are payable at the demand of the depositors (that is, whenever a depositor wants to write a check). Demand deposits (checking accounts) may, of course, originate by depositors entrusting additional cash to the bank, *but only a small fraction of total deposits have actually originated in this way. Most demand deposits*

590 THE BANKS AND THE QUANTITY OF CREDIT

have come into existence when the banks made loans or investments and the borrowers took the loans in the form of checking accounts. The manner in which the banks can and do create new credit by making loans and investments is the subject of the remainder of this chapter.

In analyzing the manner in which credit is extended by the banking system, it is helpful to refer to the balance sheet of a typical commercial bank. Such a balance sheet, for a hypothetical "Hometown Bank and Trust Company," is shown below, and it is worth while to consider briefly the makeup of this statement before proceeding.

REPORT OF CONDITION

HOMETOWN BANK AND TRUST COMPANY

Hometown, Iowa

<i>Assets</i>		<i>Liabilities</i>	
Cash	\$350,000	Deposits	
U. S. Gov't obligations . .	410,000	Demand	\$889,099
State, county and municipal bonds	250,000	Savings and time	655,504
Other bonds	110,000		<hr/>
Stock in Federal Reserve Bank	3,600		\$1,544,603
	<hr/>	Capital stock	100,000
	\$1,123,600	Surplus	20,000
Loans	550,000	Undivided profits	35,008
Overdrafts	11		
Building and fixtures	26,000		
	<hr/>		
	\$1,699,611		<hr/>
	<hr/>		\$1,699,611

Member Federal Reserve System

Member Federal Deposit Insurance Corporation

The first five assets are lumped together in this balance sheet because the bank wants to show what a large proportion of its assets are cash or almost as good as cash in case depositors should suddenly demand their funds. Usually, though not always, the assets are arranged beginning with cash and moving down to assets that are less and less easily convertible into cash on short notice. The more quickly and easily convertible into cash an asset is, the more "liquid" it is said to be. Thus United States government bonds are highly liquid because the bank can always sell them on short notice without any substantial loss in value. Building and fixtures are very illiquid because they would be difficult to sell on short notice and even if they could be sold it might well be at a considerable sacrifice in price. Obviously, not all of the range of assets held against liabilities is as liquid as cash, and commercial banks virtually never even approach such a 100 per cent cash reserve. The banker's aim is always to keep enough liquidity to meet all likely demands of his depositors.

Among the bank's liabilities, it is evident that deposits are the predominant item. Deposits are a debt, a liability, because the bank must pay cash to depositors if demanded by them. As in the case of other corporations, capital stock, surplus, and undivided profits are listed on the liability side of the balance sheet, because they represent the net worth, or "proprietaryship," of the bank, which the bank as an institution "owes" to the stockholders. Only the surplus item is substantially different from that in nonbanking corporations. When a bank is organized and stock sold, the stock is typically sold not for \$100 a share (face value) but for \$120 or \$133 a share. The extra \$20 or \$33 goes to set up the bank's surplus. Thus, if the bank's capital were to be \$100,000, as in our example, 1,000 shares of stock would be sold at \$120 per share, which would give \$100,000 for capital and an additional \$20,000 for surplus. This is simply a customary method of bank management and bookkeeping, now required by law. The net effect is the same as if \$120,000 of capital stock had been issued. The balance sheet of the Hometown Bank shows that it has not yet transferred any of its profits to increase its surplus—the surplus is merely the original \$20,000 paid in when the bank was begun. If the bank were to go out of business for any reason, all the assets would be sold and the money used first to pay the depositors, with anything left over being given to the stockholders. The capital, surplus, and undivided profits represent at any time what would be left for stockholders if all the assets of the bank were to be sold at the figures indicated and all depositors paid off. Of course if the bank had to "liquidate" in this manner, it is by no means sure that each asset would bring just the amount estimated on the balance sheet.

To illustrate the credit creation process in an oversimple example, suppose that Mr. X goes to his bank and asks to borrow \$100. After investigating X's resources, the banker decides to loan him the money; X signs a note promising to repay the \$100 at some set future date, and the bank is prepared to give him his \$100. X may take the cash, but the likelihood is that the bank will simply add \$100 to his checking account, on which he can write checks as he wants to use it. In this case the bank has received no money in addition to what it previously possessed—yet Mr. X's deposit (checking account) has been increased by \$100. The bank has "created" \$100 of new deposits, increasing by this amount the total volume of means of payment. The bank has taken X's debt (his promise to repay the \$100), which could not serve as money, and given him in exchange its own debt, which constitutes widely acceptable money. This exchange has probably benefited both parties—the bank receives interest on the deposit it loans X, while in return he obtains use of the \$100 for the period of the loan.

Against this background, Mr. X's \$100 loan from the Hometown Bank can be traced through in more detail, in order to investigate the process of bank credit creation. Mr. X takes no money but receives a \$100 credit to his checking account. The bank's balance sheet will now show \$100 larger demand deposits, while on the asset side loans have increased

\$100 since the bank has X's promise to repay the \$100.⁸ Thus the balance sheet will still balance, but with each side \$100 larger. The bank has "created" \$100 of checking deposit—the amount of cash has remained unchanged. The \$100 is a complete addition to the supply of money substitutes, and is somewhat the same as if the bank had simply printed \$100 of paper bills.

But having borrowed the money it is likely that X will spend it. Suppose he spends it all with one check a week later and the person to whom he gives the check comes and cashes the check at the bank, taking \$100 in cash. This means that two things will now happen to the bank's balance sheet. The bank will have paid out \$100 cash, so the cash item will be that much decreased. But also since X has withdrawn his deposit by writing a check, the bank's demand deposits have decreased \$100. That is, the bank has now paid the \$100 it promised to pay at X's demand. Let us compare the situation now with that before X went in to borrow the money. The bank has paid X's check—hence it has lost \$100 of cash. Deposits increased \$100 when X received the loan, but he used his new deposit by writing a check on it, which wiped out the \$100 increase, so now deposits are just what they were in the first place. But loans are still \$100 more than originally since X still owes the bank the \$100 he borrowed and spent. The balance sheet still balances (as is the most remarkable thing about balance sheets), because cash has decreased by exactly the same amount as loans have increased so that both total assets and total liabilities are the same as before the loan. No matter what sort of transaction takes place the balance sheet will always balance afterward for there will always be two or more offsetting entries.

It may clarify matters to trace through another simple transaction. Suppose that when X borrowed the \$100 he had simply taken cash. In this case, the bank's cash would have decreased \$100, its loans have increased \$100, and there would have been no effect at all on deposits. The loan would not be in the form of a checking account, so there would be no creation of credit. Money from the bank's vaults would simply be paid out, increasing the amount of money in hand-to-hand circulation. *It is only when loans are taken in the form of checking accounts that the banks can create credit.* As a matter of fact, most loans are taken in this fashion.

What happens when X repays the loan? Suppose he repays it with actual cash that he has obtained somewhere. This will increase the bank's cash by \$103 (\$100 for the loan and \$3 for interest); it will decrease loans outstanding by \$100; and it will increase profits, on the liabilities side, by \$3. The bank now has \$3 more cash than it had before

⁸ Banks ordinarily deduct interest on loans in advance. Thus the bank would give X, let us say, only \$97 and keep the other \$3 for interest; X would repay the full \$100. This process of deducting interest in advance is called "discount" rather than charging "interest." However, this detail need not concern us here and we shall assume the interest to be paid when the loan is repaid.

the loan was made, and its undivided profits have increased \$3, so that both sides of the balance sheet are \$3 larger than before. Otherwise the bank is back in its original condition. But in the meantime it has enabled Mr. X to spend \$100 (by writing a check), thus effectively increasing the quantity of credit during this period by \$100.

Credit creation of exactly the same sort may take place through the purchase of bonds or other securities by the bank. Suppose that instead of loaning X \$100 the bank buys a U. S. government bond and pays for it by giving the federal government a \$100 checking account. Again credit has been created, except that this time it is the government which has the power to write the checks. In all these cases there has been a \$100 increase in the means of payment without any increase in the actual money in existence.

Limits to credit creation by an individual bank

Why do not banks simply keep on expanding credit and earning interest on larger and larger amounts of loans and investments, if all they have to do is create new checking accounts? The answer is, of course, that there is always the danger that the depositors will ask for cash or write checks that have to be paid. Hence, if any banker were to expand his loans and investments very greatly by creating new credit he might be in a very tight spot if a large number of these new deposit owners were to ask for cash at the same time. The reason he is at all safe in holding deposits beyond the cash he has is that it is very unlikely that all his many depositors will ask for cash or write checks at the same time. But if he were to expand credit to, let us say, a hundred times the amount of cash he has, the demand of a very small proportion of his depositors for cash could exhaust his cash and force him to close the bank (in which case we should say that the bank had "gone broke," as did so many banks during the recent depression when they were unable to get enough cash to meet the sudden cash demands upon them). The possible demands of depositors or recipients of depositors' checks for cash exercise a very stringent restraint on the amount of credit a banker may safely extend through making new loans and investments.

That bankers must keep certain amounts of cash in relation to the amount of deposits they have is required by law; each bank is required to have cash "reserves" equal to a certain per cent of its deposits. This is what is meant by "bank reserves." The money that constitutes the bank's reserve is not usually kept in its vaults. Instead the bank's reserves are usually held by one of the twelve Federal Reserve Banks, but this is a detail that need not concern us at present. A bank can always get its reserve from the Federal Reserve Bank on a moment's notice if the reserve is needed to meet depositors' demands.

Although reserve requirements for state banks vary somewhat according to state laws, for the greater part of the banking system the following reserve requirements were prescribed in late 1942. For all savings and time deposits, each bank was required to keep a reserve equal to 6 per

cent of the deposits. This reserve is very low—supposedly savers are very unlikely to withdraw savings or time deposits on short notice, so that there is little necessity for the banks to have much cash reserve against them. If, however, people do demand cash for their savings and time deposits in considerable numbers (as has been the case during “runs” on banks in periods of financial crisis), the banks are likely to find their cash reserves exhausted very soon, and they hesitate to enforce thirty- or sixty-day withdrawal notice provisions, which might be interpreted as signs of weakness. Because banks pay interest on savings and time deposits (usually 1 per cent or 2 per cent at present) but pay no interest on demand deposits, there is a special incentive to depositors to put their funds in savings rather than demand deposits whenever they do not want to write checks on the funds deposited, even though the funds are not long-period savings. The effect of this distinction between types of deposits is to lower substantially the required reserves for any given volume of deposits.

The reserves that must be held behind demand deposits vary according to the size of the city in which the bank is situated. Banks in New York and Chicago must hold 20 per cent reserves; banks in “reserve cities” (roughly, cities smaller than Chicago but above 100,000 population) must also now hold 20 per cent reserves; and banks in all smaller cities must hold 14 per cent. These requirements are higher than those on savings deposits because banks must be ready to pay out demand deposits on a moment’s notice, which is not required in the case of savings deposits. Reserve requirements effectively set an upper limit on the amount of credit a bank can expand on a given amount of cash.⁴ Since the Hometown Bank in our example is in the smallest class town, its reserve requirement on demand deposits would be 14 per cent. This means that the Hometown Bank must always keep at least 14 cents in cash reserve for every dollar of demand deposits it has. If, then, it had \$140,000 of cash reserve, the largest amount to which it could legally expand its demand deposits by making loans and investments would be \$1,000,000. If its cash reserve were increased to \$210,000, this would make it legally possible to expand its demand deposits up to roughly \$1,500,000 (because \$210,000 is 14 per cent of \$1,500,000). If its cash reserve were decreased, this would lower the maximum volume of deposits that it might hold.

There is always a pressure on the banker to extend as much credit as he thinks is safely possible, for he obtains no interest on his money while it is lying idle as reserve. Even though the rate obtained on loans or investments is low, it is advantageous to keep excess reserves at the lowest “safe” level. The major check on this expansion of loans and investments on the basis of available excess reserves is, as previously

⁴ Reserves thus have another function, in addition to protecting depositors. By controlling the amount of bank reserves, the Federal Reserve and Federal Treasury have a direct means of influencing the amount of bank credit extended. This very important function of reserves as a control device is emphasized later.

noted, the danger of loss of the reserves through cash withdrawals by depositors or through presentation of depositors' checks for payment. In order to provide a safe margin above required reserves, bankers must keep excess reserves sufficient to meet ordinary and unexpected possible cash withdrawals without impairing the legally required reserves. Though bankers ordinarily can estimate fairly closely what the demands for cash will be at any given season or time of the week, conservative banking practice demands that excess reserves somewhat above actual expected needs be kept. But the *most important limitation on the power of an individual bank to expand credit to the legal limit on excess reserves is the fear of "adverse clearing balances."*

A simple example will indicate the significance of adverse clearing balances. Suppose that the Hometown Bank has \$1,000 in excess reserves, and that the Hometown banker decides to make loans on the excess reserves. He could legally expand loans and investments to \$7,143 on this excess reserve, assuming a 14 per cent reserve requirement. (Since \$1,000 is 14 per cent of \$7,143, deposits could be increased by \$7,143 and still remain within the legal requirement.) But if he were to begin such an expansion of loans and deposits, certainly some of his borrowers would write checks payable to persons who did not do business with the Hometown Bank. Suppose a loan is made to X in Hometown, who writes a \$100 check to Y in Chicago for new farm machinery. Y cashes the check at his bank in Chicago, and the Chicago bank immediately sends the check to the Hometown Bank to be redeemed in cash. The Hometown Bank deducts \$100 from X's account, sends the \$100 to the Chicago bank, which has already paid \$100 out to Y; and the transaction is completed.⁵ The Hometown Bank has lost \$100 cash to the Chicago bank; \$100 of the original \$1,000 excess reserve has been lost by having a single small depositor's check sent in by another bank to be paid in cash. And it is likely that many of those persons to whom the Hometown Bank has made new loans will send checks to persons banking at other institutions and the checks will have to be paid in cash. Therefore, the Hometown Bank would lose all its excess reserves in this manner long before it had expanded credit anywhere near \$7,143. *The fact that checks are apt to be written on it and presented for cash payment by other banks puts a very sharp restriction on the amount of credit that any individual bank can expand on any given excess reserve.* Thus any bank expanding its loans and investments is likely to lose reserves because of "adverse clearing balances" as other banks present more checks that it must pay them than it has to present to the other banks for them to pay it.

Ordinarily the checks written on any bank and presented to it for payment by other banks will roughly equal the checks it has to collect from other banks. Thus, the Hometown Bank will ordinarily owe other banks

⁵ The process by which checks are collected and "cleared" between banks is discussed in more detail later. Actually little cash changes hands.

for checks written on it amounts roughly equal to the amounts due it because of checks it receives for collection. For example, X has sent a \$100 check to Chicago for his machinery, but a Chicago packing plant may send a check for \$100 to a Hometown farmer for livestock, which the farmer cashes or deposits at the Hometown Bank. The Hometown Bank owes the Chicago bank \$100 on X's check, but this is offset by the \$100 the Chicago bank owes on the livestock check. Although ordinarily these two types of checks will roughly offset each other, if the Hometown Bank expands its loans sharply while other banks do not, it becomes likely that checks written on the Hometown Bank will increase without a corresponding increase at other banks. *Recognizing this likelihood, no banker would ever attempt to make new loans and investments anywhere near \$7,143 on \$1,000 of excess reserves unless he had some special reason to suppose that he would not lose reserves through adverse clearing balances. Indeed, he will hesitate to extend new credit much beyond his excess reserves, in this case much beyond \$1,000.*

What has been said thus far about banks and the quantity of credit may be recapitulated briefly: (1) one function of banks is to convert savings into investment, which function has no direct effect on the quantity of credit; (2) commercial banks are distinguished from savings banks and other such institutions in that commercial banks as a group do not simply loan out money that people deposit, but actually "create" credit; (3) legal reserve requirements set by federal or state authority definitely limit the amount of credit a commercial bank can expand on any given supply of reserves; and (4) the power of an individual commercial bank to expand credit even when it has excess reserves is sharply limited by the danger of adverse clearing balances and of cash withdrawals into hand-to-hand circulation.

Credit creation by the banking system

Any one bank attempting to expand loans and investments much beyond its excess reserves when other banks are not expanding is sharply checked by the factors noted above, especially the likelihood of adverse clearing balances. But when the banking system is viewed as a whole, this limitation of adverse clearing balances disappears. Let us assume again that the Hometown Bank has \$1,000 excess reserves, and that all other banks are loaned up to their legal limits. As the Hometown Bank begins to make new loans and investments, its reserves are gradually drawn away to other banks, limiting its credit expansion possibilities. But what the Hometown Bank loses, some other bank gains. Suppose the Hometown Bank makes a new loan of \$100 (reducing its excess reserves by \$14 to \$986), and the borrower writes a check that is deposited in another bank. When this check is returned for collection, the Hometown Bank's excess reserves are drawn down to \$900, while the other bank has gained \$100 of reserves and deposits, giving it \$86 of excess reserves. The total bank credit extended has increased by \$100, and the total excess reserves have decreased by \$14, to \$986, part of the excess being in each of

the two banks. Each bank is now in a position to expand loans and investments further on the basis of its excess reserves, but as each does so it will lose reserves to other banks through adverse clearing balances.

As this process continues, the original \$1,000 excess reserves of the Hometown Bank may get so spread out that each of 100 banks has \$10 of it, or each of 1,000 banks \$1 of it. But what one bank loses another gains, and for the banking system as a whole there is no loss of reserves through adverse clearing balances. Therefore, adverse clearing balances do not make it impossible for all banks *taken as a group* to expand credit to the legal limit. The \$1,000 of excess reserves would not permit one bank acting alone to expand credit much beyond \$1,000, but the banking system as a whole may expand to the full \$7,143 because of the absence of the limitation imposed by adverse clearing balances.

Actually, most of the important cases of bank credit expansion have come when most banks are expanding loans and investments more or less simultaneously. Especially has this been true in the revival and boom phases of business cycles, when the expansion of bank credit plays a central part in increasing money demand for goods, which in turn tends to bring increased production and employment and higher prices. But even when the banking system as a whole is expanding in such cases, it is subject to the other limitation noted—namely, cash withdrawals into hand-to-hand circulation. When cash is withdrawn from any bank this decreases the total reserves of the banking system, and hence limits not only the expansion possibilities of the paying bank but also the expansion possibilities of the system as a whole. In good times such a cash withdrawal is likely to take place as business improves, thus limiting the expansion powers of the system below the legal limit of \$7,143 on an original \$1,000 reserve, and this limitation is reinforced by the traditional banking standards that demand a “safe” amount of excess reserves beyond expected needs. But the expansion powers of the banking system with banks expanding simultaneously are great and may account for tremendous increases in the money demand for goods and services. In crisis periods, runs on banks may account for large cash withdrawals, which similarly reduces the amount of credit the banking system can support.

Contraction of credit by commercial banks

Just as commercial banks can and do expand or “create” credit through making loans and investments, so they can and do contract or “destroy” credit through calling in loans and investments. Thus if the Hometown Bank for some reason (let us say, fear as to the safety of the loans) decides to collect all loans coming due and to refuse to make any new loans, this would effectively decrease the total quantity of bank credit (deposits), just as making the new loans increased it. Moreover, if banks with small excess reserves have heavy cash withdrawals and are forced to call in loans to obtain the cash, every dollar of cash reserve lost necessitates the contraction of several dollars of loans and investments. For example, if one dollar of reserves is the basis for five dollars

of outstanding loans and investments, then loss of each dollar of reserves forces contraction of five dollars of loans and investments.

Just as the important cases of credit expansion have ordinarily come when many or all banks expand together, so the important cases of contraction have come when many or all banks are simultaneously calling in loans and liquidating investments. The disastrous results that attend a mass contraction of bank credit are clearly indicated by the period following 1929. Bankers became alarmed and feared that customers would be unable to repay loans because of the deepening depression; hence they endeavored to collect loans as rapidly as possible, refusing to make new advances. Many securities were also rapidly liquidated by sale on the open market. Simultaneously people were withdrawing cash from the banks to hoard in safe deposit boxes, mattresses, old shoes, and wherenot, which forced the banks to call still more loans and to liquidate investments to get cash to meet withdrawals. By calling for payment of loans, banks forced borrowers to sell securities and stocks of goods at forced sale, just when deepening depression and pessimism made people most hesitant to buy and when the banks also were selling securities. Thus funds to repay bank loans could be obtained only by taking large losses on securities and goods sold. And the greater the losses sustained, the more worried bankers became about the security of remaining outstanding loans and the harder they pressed for repayment, reinforcing the downward spiral of deflation and liquidation. Thus the process of credit contraction, once well started in a time of uncertainty and deepening depression, is likely to force bankers to contract credit drastically in order to protect the solvency of the banks, a variety of contraction that strongly reinforces the cumulative downward spiral of deflation. This is only one small segment of the process of deepening depression, but it shows partially the role that may be played by bank credit contraction in the downward spiral. Fuller discussion of the part played by the banking system in business cycles will be undertaken in considering booms and depressions.

Critics have emphasized that banks act like little individual "mints," coining and destroying credit just as the federal mints coin and destroy actual money. This is not properly a criticism of bankers themselves, but rather of the banking system, which leads them to expand loans and investments greatly during boom times in the search for maximum profits and which forces them to contract loans and investments drastically in crisis periods in an attempt to maintain the solvency of their banks—even though this expansion and contraction lead to much more severe booms and depressions than would otherwise occur. The fact that banks are required to hold only small fractional reserves behind deposits makes it possible for them greatly to expand loans and investments on relatively small reserves, and the desire for profit leads them to take advantage of this expansion possibility. The fact that expansion has gone so far forces drastic contraction in bad times when depression sets in. This creation and destruction of credit are possible only because banks do not have to

hold 100 per cent reserves behind deposits. And the lower the percentage of reserves required, the greater becomes the expansion possibility in good times and the likelihood of very drastic contraction in bad times.⁶

⁶ In view of the important part bank credit expansion and contraction play in business cycles, some critics advocate requiring 100 per cent reserves behind all bank deposits, thus mitigating the reinforcing part bank credit plays in booms and depressions. It is argued that this step would do much toward lessening the severity of business fluctuations. And its advocates point out that businessmen and others wishing to borrow could continue to do so, through "savings and loan organizations" that would receive funds that savers wanted to invest and loan these savings to businessmen or others wanting to use them. Such organizations could loan out only as much as savers deposited, rather than being able to create credit as the banks do at present. The present banks would presumably be split into two parts: one part to act as a storehouse for funds that savers simply wanted to keep idle for checking or other purposes, on which depositors would probably have to pay storage charges; the other part to act as a savings and loan organization to receive savings, paying interest on them, and to loan out these savings at a somewhat higher rate of interest in order to obtain funds to pay expenses, interest to savers, and have some profit left over.

Undoubtedly the plan has merit, but there are important difficulties involved; perhaps most important are that of transition to such an arrangement, and that of other "money substitutes" that might arise. The proposal and the problems it raises are too complex for detailed discussion here. It is important to see that by increasing the reserve requirements behind deposits we should eliminate more and more of the part that bank credit plays in making business fluctuations so severe. But it is equally important to recognize that the 100 per cent reserve plan would *not* by itself be anything approaching a "cure" for general business fluctuations.

CHAPTER 40

The Price Level; or the Value of Money

What Is Meant by the "Value of Money"?

WE USUALLY state the values of things in dollars and cents. Using money as a standard or common denominator of value is a more convenient method than stating that the value of a radio is a sack of flour plus a sweater plus half a dozen eggs plus a lipstick plus a pork chop plus the rest of the group of goods and services that the radio could be exchanged for. Thus when we say the radio is worth \$50, we are really saying that it is worth any group of goods and services which can be obtained for \$50. The money is merely a common denominator—a short-cut for listing all the innumerable combinations of goods and services that can be obtained for \$50. The study of individual prices in traditional "price and distribution theory" is devoted largely to determining why the value of any particular commodity or service is what it is *relative* to other goods and services. For example, this means explaining why one orange will be worth two apples, even though we say that one orange will be worth a nickel while apples are two for a nickel. So far as price and distribution theory is concerned, the prices could just as well be 10 cents and 5 cents or \$2 and \$1—without specific consideration of money as such, economic analysis is concerned only with relative, not absolute, prices.

Suppose the price of oranges goes up from 5 cents to 7 cents apiece. This may mean that the 7 cents that you must now spend to buy an orange requires giving up more of other things than were foregone in spending 5 cents to get an orange before. That is, other prices may have remained unchanged while that of oranges varied. But other prices may have risen too. Suppose that for some reason the price of apples and of bananas has gone up in just the same proportion as that of oranges. Then, although the price of oranges has risen, the relative value of oranges as compared with apples and bananas is precisely what it was before the price rise. And if we suppose that the prices of *all* goods and services rise in exactly the same proportion as the price of oranges, the relative value of every good and service would be absolutely unchanged although the prices of all of them would be different—higher by some 20 per cent. In this case, while the relative values of all other things have remained unchanged, the value of money (of the common denominator) has changed. One dollar will no longer buy as many goods and services as it would before. The value of

the dollar has fallen relative to the value of all other things, just as in the first case the value of the orange had risen relative to the value of all other things. Fewer goods and services can be obtained for one dollar; it takes more dollars to get the same goods.

It is more common to speak of changes in the price level than of changes in the value of money. In the above example, the value of money has fallen and the price level (average of all prices) has risen. The two are simply different ways of saying the same thing. The value of money varies *inversely* with the price level. To use a more realistic example, in 1929 an income of \$1,500 would have been sufficient to obtain a certain group of goods and services equal roughly to what was called by some a "subsistence standard of living." In 1933, on the other hand, this iden-

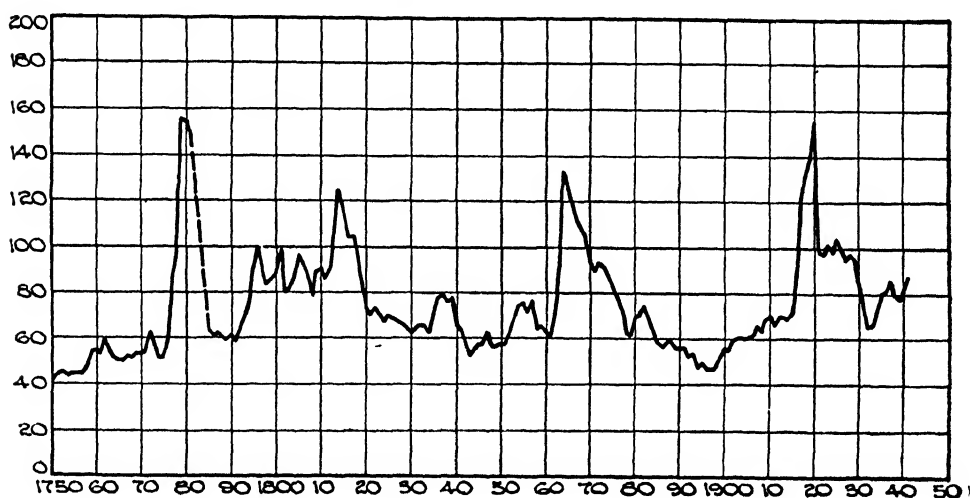


Fig. 40—1. Wholesale price level, 1750–1941 (1926 = 100).*

* Warren and Pearson index to 1890, from *Wholesale Prices, 1931*, U. S. Bureau of Labor Statistics, pp. 112–114; U. S. Bureau of Labor Statistics index since 1890.

tical collection of goods and services could have been obtained with an income of around \$900. That is, the price level fell very sharply in the depression following 1929, which meant that the value, or purchasing power, of each dollar was much greater in 1933 than in 1929. (It hardly needs be pointed out that it does not follow that people were therefore better off. The price level fell sharply, but incomes also fell.)

Figure 40—1 indicates roughly the great fluctuations of the wholesale price level from 1750 to 1941. The very large fluctuations show clearly that money (the dollar) has not been a stable standard of value. On the contrary, money has proved to be a very unstable common denominator. It is as if the yardstick had intermittently stretched to 6 feet and then contracted to 18 inches. This instability is not necessarily inherent in money—it is partly a result of the nature of money and credit, partly a result of bad management of that money and credit, and partly a result of nonmonetary forces.

Price Indexes and Price Level Changes

In the illustrative example just presented it was assumed that all prices rose in the same proportion, the value of money falling in the same degree. It is almost inconceivable that this should actually happen. Even in such a sharp price decline as that following 1929 not only did some prices drop much faster and farther than others, but a few even rose slightly for special reasons (for example, anhydrous ammonia, for which new uses were found). Yet, though not all prices fell, we say the price level fell because there was a fall in the average of all prices. Four hypothetical prices may be taken as an example to illustrate the means of telling whether their average went up or down between 1929 and 1933 and how much.

TABLE 40—1

HYPOTHETICAL PRICE LEVEL FOR 1933 WITH 1929 AS BASE YEAR

	1929	1933
Cotton, per lb.	6 cents = 100%	2 cents = 33% (of the 1929 price)
Peanuts, per lb.	15 cents = 100%	10 cents = 67% (" " " ")
Ammonia, per pint . . .	50 cents = 100%	55 cents = 110% (" " " ")
Apples, per lb.	6 cents = 100%	3 cents = 50% (" " " ")
	4)400%	4)260%
	100% — price level in 1929	65% — price level in 1933

Using 1929 as the "base year," the price of each item in that year is 100 per cent of itself; this is what is meant when it is said that "1929 is taken as 100." Then the percentage that each 1933 price is of the 1929 price of the same commodity is computed. For example, the 2-cent price of cotton is only 33 per cent of the 1929 price. This 33 per cent is the 1933 "price relative," since it shows the relation of the 1933 to the 1929 price. To find the average amount prices have declined (the decline in the price level), the four price relatives for 1933 are averaged by adding them together and dividing by four. Thus the 1933 average is 65 per cent; put in other words, the price level has declined from 100 in 1929 to 65 in 1933. (The per cent is usually omitted for convenience.) The 100 in 1929 and the 65 in 1933 are called "index numbers" or indexes of the price level. Of course most price indexes include more commodities, but the process behind them is similar.

This very simple method of calculating price level changes is likely to give misleading results, however. It tacitly assumes that cotton, peanuts, ammonia, and apples are of equal importance. That is, a 10 per cent change in the price of peanuts influences the index exactly as much as a 10 per cent change in the price of cotton or a 10 per cent change in the price of apples. Actually cotton is tremendously more important in the economy than peanuts or apples, and it seems logical that a change in the price of cotton should affect the index more than minor products. To give added importance in the index to more important commodities, each price

is usually "weighted" according to its relative importance among the prices being used—that is, more important commodities are given greater weight than the less important ones. The commonest method is to weight each commodity in proportion to its importance in the total expenditures involved. For example, suppose 300,000,000 pounds of cotton were sold in 1933 at 2 cents per pound, then total expenditures on cotton that year would be \$6,000,000. Suppose 1,500,000 pounds of peanuts were sold at 10 cents per pound, then total expenditures on peanuts would be \$150,000. Suppose 100,000 pints of ammonia were sold at 55 cents per pint, the total expenditures on ammonia would be \$55,000. And suppose 20,000,000 pounds of apples were sold at 3 cents per pound, the total expenditures on apples would be \$600,000. Total expenditures on the whole list would then be \$6,805,000. The respective weight of each of the four com-

TABLE 40—2
HYPOTHETICAL WEIGHTED PRICE LEVEL FOR 1933 AND 1935
WITH 1929 AS BASE YEAR

1933					1935				
	Price	Relative	Weight	R × W		Price	Relative	Weight	R × W
Cotton, lb.	2¢	33	88	2904	Cotton, lb. . .	4¢	67	86	5762
Peanuts, lb. .	10¢	67	2	134	Peanuts, lb. .	12¢	80	3	240
Ammonia, pt.	55¢	110	1	110	Ammonia, pt.	55¢	110	2	220
Apples, lb.	3¢	50	9	450	Apples, lb. . .	4¢	67	9	603
			100	3598				100	6825
				.359					.68
				or					or
1933 weighted price level—36%					1935 weighted price level—68%				

modities would be the per cent it is of the total expenditures. Thus, cotton would have a weight of 88, peanuts of 2 +, ammonia of 1 —, and apples of 9, out of a total weight of 100.

Having obtained the weights to be used for each commodity in 1933 the weighted price level for 1933 can be obtained, still taking 1929 as the base (100 per cent). First, as before, the relative price of each commodity in 1933 as compared with 1929 is computed. For example, cotton was 6 cents in 1929, 2 cents in 1933—the relative for 1933 is 33 per cent. Second, each 1933 price relative is multiplied by its weight. For example, the relative of 33 for cotton is multiplied by its weight of 88, getting 2,904, and the same process is followed for each of the other commodities. Third, these results are added for all four commodities, and divided by the sum of the weights (that is, by 100). The quotient is the 1933 weighted price index of 36 per cent.

This process is illustrated in Table 40—2, which gives the 1933 price, the 1933 price relative to 1929 price, the 1933 weight, and the result of multiplying the relative by the weight, for each commodity, and the totals necessary to obtain the 1933 weighted index. An index for 1935 is calculated in the same way to show how the results might look as prices again began to rise. The 1935 price relative is also computed on the 1929

base for each commodity, and then multiplied by its weight. As indicated by the table, it is assumed that the relative importance of cotton was somewhat decreased in 1935 as compared with 1933, with corresponding increases in the relative importance of other commodities. Using the altered weights and the prices shown, the 1935 index is found to be 68.

In comparing the weighted price index for 1933 (Table 40—2) with the unweighted index for that year (Table 40—1), it becomes evident that the weighted index shows a much greater decline from 1929. This is because it gives more influence to cotton, which fell very sharply in price, and less influence to minor commodities, which either fell less or even rose in price. Although this hypothetical example is much too small to represent reality, it does show the method used if a real price level index

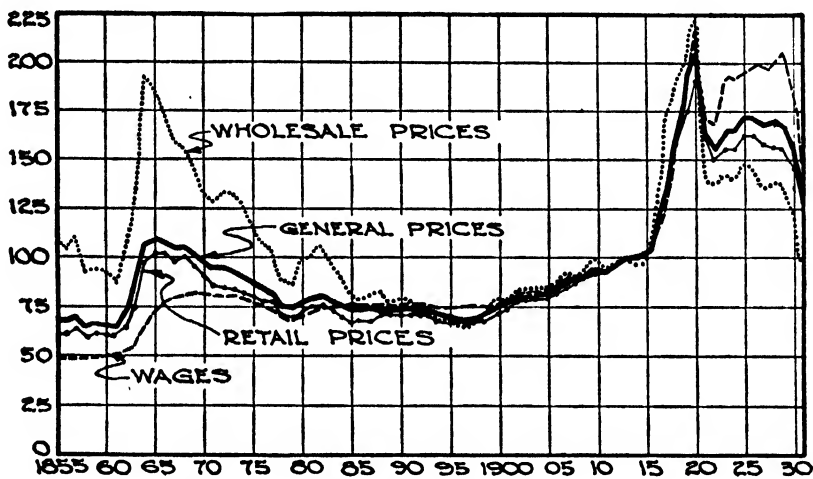


Fig. 40—2. Varying price levels, 1855—1932 (1913 = 100).*

* Reproduced from F. Cyril James, *The Economics of Money, Credit and Banking* (Ronald Press, New York City, 2nd ed., 1935), p. 515, by permission of the copyright owner.

is to be derived, and indicates the more meaningful results given by a weighted index. For example, the United States Bureau of Labor Statistics uses 784 different prices in calculating its index of wholesale prices, but it uses this weighted method. Virtually everyone agrees that weighted index numbers must be used if meaningful results are to be obtained.

So far it has been tacitly assumed that there is just one price level ("the" price level) and one price index. But one may equally well talk about a number of different price levels—really different segments of a general price level. Thus there may be indexes of wholesale prices, or of retail prices, or of the cost of living, or of some other group of prices. Although all of these would be likely to give somewhat similar results over any given period, they would usually diverge somewhat, possibly a good deal.

Figure 40—2 shows the movements of various United States price levels over three fourths of a century. What index is best to use in any case

depends primarily on what one wants to use it for. If the purpose is to measure changes in the cost of living for any group, the logical choice is an index composed of prices of the goods and services consumed by the group in question—for example, retail prices on food, clothing, furniture, rent, light, heat, and so on. But if the purpose is to study the business cycle, an index of wholesale prices is much more useful because it fluctuates more quickly and widely in response to changed business conditions. Other purposes would call for other indexes.

By “the” price level is usually meant an average of *all* prices, and it is this general average that is usually meant when “the value of money” is discussed, even though the “value of money” to any one person or group might be more accurately measured by a price index including only prices of those goods and services that the person or group buys. An examination of the elements involved in determining “the value of money” and changes in its value not only helps to provide an explanation of changes in “the” price level, but the procedure used is also applicable to any particular price level which may be of interest. Actually, it is the fact that different prices and groups of prices change at very different rates that makes price level fluctuations of such great significance, as will become increasingly evident.

Price Level Changes and the Distribution of Income

Not only are there many different price levels, but within each price level there is likely to be a considerable diversity of change among the various individual prices included. When price levels are changing, some whole groups of prices move more rapidly than others. For example, retail price changes usually lag behind wholesale, and finished goods prices behind raw material prices. The wages of nonunion workers typically rise more slowly than wholesale prices and the cost of living, while the wages of unionized workers are likely to rise more rapidly. Some contracts are made for long periods ahead and hold to their original prices regardless of price level changes. Rates of return on different kinds of investments vary greatly with changing price levels.

The fact that various prices change at different rates, especially in booms and depressions, means that some persons gain relative to others as prices change. To take a very simple example, if a teacher has a three-year contract at \$2,000 and the prices comprising his cost-of-living rise, his real income is decreased. Although he still receives the same money income (price for his services), the purchasing power of this income is decreased by the higher prices he must pay for what he buys. This simple example is indicative of the general principle *that changing price levels involve special gains and losses to various groups largely because of the fact that some prices change more rapidly than others.*

Table 40—3 gives a rough indication of the relative positions of persons receiving incomes from different sources in 1929, 1932, and 1935, a period of very sharp price level changes. All figures in the table are percentages,

showing the relation of 1932 and 1935 to 1929. The first horizontal row of figures shows the relation of the national money income in 1929 to that in 1932 and 1935. The second row shows the movement of the cost of living over the same years. Dividing the cost of living into the money incomes people received gives the purchasing power of incomes over the items included in the cost of living index. This purchasing power of money incomes is shown in the lower portion of the table, first for all incomes, and then for the income received from various functional sources. The table shows that the total income received as dividends suffered relatively the greatest drop in purchasing power, while interest income actually rose considerably in purchasing power. Interest payments fell

TABLE 40—3

PURCHASING POWER OF INCOMES FROM DIFFERENT SOURCES (1929-1935) *

	1929	1932	1935
Money Incomes as a Whole	100%	62%	68%
Cost of Living Index,** by which Money Incomes are Converted into Purchasing Power	100	80	81

PURCHASING POWER

All Incomes (except compensation for injuries, pensions, and relief)	100	78	84
Wages and Salaries	100	76	86
Interest	100	124	101
Dividends	100	58	60
Entrepreneurial Incomes	100	66	73

* Income figures from National Bureau of Economic Research Bulletin No. 66 (1937), Table IV. The income figures were first related to 1929 as a base year; and then in the lower half of the table converted into purchasing power figures by dividing by the cost of living index.

** Cost of living index of the U. S. Bureau of Labor Statistics.

relatively little between 1929 and 1932, while the cost of living fell rapidly, so the purchasing power of interest income rose substantially. The term "entrepreneurial incomes" is of course a heterogeneous combination of profits, wages of management, and interest on investment.

Although the table shows very roughly what happened to the purchasing power of persons receiving income from these various sources as the economic system passed from prosperity to deep depression to partial recovery, it is important to remember that these figures show only the purchasing power of broad functional income groups. Within these groups, various individuals may have been very differently affected. For example, it is slight consolation to a bondholder whose interest has been defaulted to know that the purchasing power of interest-receivers as a group increased during the depression. Nor was it much consolation to an unemployed man to know that the purchasing power of wages and salaries as a group increased from 1932 to 1935. On the other hand, workers whose wages or salaries remained fixed during the depression found their purchasing power sharply increased (assuming that they remained employed), even though the purchasing power of the wage and

salary group dropped sharply from 1929 to 1932. In addition to the fact that there were great differential effects within groups, it must be remembered that large numbers of people receive incomes from more than one source so that for any individual it may be necessary to combine two or more of the income indexes.

From these considerations it should be clear that one cannot generalize very accurately about the effects of changing price levels on broad occupational or institutional groups. The following paragraphs attempt only to set forth a few of the most basic conclusions that may be drawn. Since except for war periods the sharpest price level changes come in booms and depressions, and since in the slower price level changes it is much easier for slowly changing incomes to be adjusted to changing prices, it is advantageous to approach the problem primarily through price level changes in booms and depressions, even though a complete discussion of these effects is impossible until after analysis of general business fluctuations.

The basic underlying proposition is simple: If one's income lags behind rising prices for the things he buys, his purchasing power is decreased; if one's income lags behind falling prices for the things he buys, his purchasing power is increased; if one's income changes either way at the same rate as prices of the things he buys, his purchasing power is unchanged. One of the simplest cases, that of persons receiving incomes fixed in money terms over rather long periods, may be taken to illustrate the application of this principle. Take for example an elderly gentleman living on the interest he receives from savings invested in fixed-interest bonds. If the price level of things he buys goes up, the purchasing power of the fixed interest receipts decreases proportionately. Fixed-interest bonds are likely to be a relatively bad investment in periods of rising prices. The holder of such securities suffers greatly from inflation, and there are many individuals and institutions (such as schools, hospitals, and charitable foundations) in this bondholding group. On the other hand, it is commonly stated that bondholders gain from falling prices, since with lower prices their interest receipts will buy more goods and services than previously. This is so *if* the bond interest is received, but since sharply falling prices are usually associated with business depression many firms are unable to make interest payments in such periods. Holders of high-grade bonds are therefore likely to gain in purchasing power with falling prices; holders of low-grade bonds may well lose all or part of their interest income through default.

The position of creditors

These statements about the position of bondholders under rising or falling prices are applicable to creditors in general. If the creditor's income is relatively fixed in money terms, he loses from rising prices of the things he buys, and he gains from falling prices *if* his income is paid in spite of the falling prices and worsened business conditions that usually accompany them. This may be illustrated by a salesman or clerical

worker who is paid a definite yearly salary. If prices rise while his salary remains unchanged, his purchasing power is decreased; if prices fall while his salary remains unchanged, and he keeps his job, his purchasing power is increased. If he loses his job, obviously his purchasing power is decreased in spite of the lower prices. The main difference between the salaried employee and the bondholder is this: the income of the bondholder is fixed over the life of the bond, while the salary of the employee is likely to be adjusted within a year or two if prices and business conditions change drastically. Since this adjustment in salaries ordinarily takes place rather slowly, salaried employees' incomes usually lag behind price level changes and salaried persons are therefore temporarily in the same position as bondholders. Their purchasing power is likely to fall at least temporarily during price level increases and rise temporarily during price level declines if they are fortunate enough to keep their jobs.

It has long been a standard textbook proposition that wage earners are in a position similar to that of salaried employees except that wages are adjusted somewhat more rapidly than salaries. Therefore it is commonly said that wage earners lose purchasing power with rising prices and gain it with falling prices. This statement is at best a half-truth as applied to recent years. There are two major modifications to be made. First, even though wage rates lag behind rising prices the real incomes of many workers rise because they get more work with higher prices and better business conditions. Conversely, even though wage rates lag behind falling prices, the real incomes of many workers decrease because they get less work with lower prices and worsened business conditions. Second, it is no longer accurate to generalize that wages lag behind rising prices since now organized labor has become so strong in many fields. For example, in numerous major fields wage rates rose more rapidly than prices during the recovery following 1933. By comparison, the wage rates of unorganized workers lag more behind rising prices and lag less behind falling prices—that is, they rise more slowly and fall faster. Whether the unorganized workers as a group are therefore worse off than organized groups with rising or falling prices depends on whether the increased employment they receive because of lower wage rates more or less than offsets the lower wage rates. With good times and rising prices it is generally safe to say that unorganized workers are the worse off in terms of purchasing power because they probably would have been hired even at higher wage rates. With bad times and falling prices it is an open question whether the high wage rates maintained by particular unions, resulting in less employment, give these unionized workers higher real incomes than those of workers with lower wage rates but more work; those union workers who keep their jobs at the high wage rates are better off, but taking the groups as wholes the question is not easy to answer. In large part the conclusion depends on the monetary policy pursued by the government in checking deflation. Wage rates held high in particular industries while product prices fall result in unemployment. But if the

government takes steps to check the decline in prices and to maintain investment activity, high wage rates may not mean unemployment. Until the problem of business fluctuations and economic policy to alleviate them has been considered, it is impossible to say more than that the effect of changing price levels on wage earners varies greatly from group to group depending on the circumstances of each situation.

The position of debtors

Debtors owing sums relatively fixed in money terms (for example, individuals who have borrowed money or corporations paying interest on bonds) gain from higher prices because each dollar paid out means less purchasing power given up. They are in an unfavorable position when prices fall since each dollar paid represents a greater sacrifice of purchasing power. Moreover, if bad business conditions force them into default and bankruptcy both they and their creditors lose.

Entrepreneurs are in a sense the largest debtors in society because it is they who pay out incomes to owners of productive agents, although of course there are many nonentrepreneurs who are debtors. To the extent that the payments made for productive services (that is, costs) lag behind rising selling prices, entrepreneurs make increased profits. To the extent that payments made for productive services (that is, costs) lag behind falling selling prices, entrepreneurs suffer increased losses. As a matter of observed fact, most payments made for productive services do lag behind selling prices, in both rising and falling price periods. Therefore, ordinarily entrepreneurs make profits with rising price levels and suffer losses (or make smaller profits) with falling price levels. Indeed, this is one of the essential reasons booms and depressions continue, once started. With falling prices but lagging costs, losses are suffered, resulting in curtailed production, employment, and workers' incomes—these curtailments in turn lead to decreased spending and still lower prices, and so on in a downward spiral. Roughly the reverse is true in a boom. The matter of entrepreneurial profits is closely dependent on business conditions and changing price levels, and all three are so closely interdependent that discussion of any one in isolation is unreal.

Since stockholders' incomes (dividends) depend largely on the profits made by corporations, stockholders' incomes tend to fluctuate directly with changing price levels and business conditions. Therefore, common stock is often considered a good buy in periods of rising prices since higher dividends are likely to be forthcoming, a bad buy in periods of falling prices since dividends are likely to be decreased. The stockholder reaps large dividends in good times; in depression he is apt to go empty-handed. However, with both stocks and bonds it is necessary to know not only expected returns (interest or dividends) but also expected security prices in relation to present prices in order to determine the "best" investment. For example, even though a stock selling for \$100 is likely to pay only \$5 dividends, while a bond at the same price may

pay \$6, if the price of the stock now seems likely to go up to \$150 with higher prices and better business, the stock would be a "better buy." Conversely, even though the stock now pays \$8 a year and the bond only \$6, if the price of the stock is now likely to fall with falling prices and worsening business, the bond may nevertheless be the "better buy." Thus in booms and rising prices many people buy common stocks both because of expected dividends and expected rises in stock prices, and the resulting demand for stocks is a primary factor leading to realization of the expected higher prices. In depressions, people tend to sell stocks both because of expected lower dividends and expected declines in stock prices, and this very selling of stocks is a primary factor leading to realization of the expected lower prices. On the other hand, the tendency is to sell high-grade bonds during booms and rising prices, and to buy high-grade bonds during falling prices and depressions, because of the different nature of the return and risk on high-grade bonds. Here it again should be apparent that we cannot meaningfully talk simply of price changes, but must bring in also the changing business conditions that ordinarily come with changing price levels.

Lastly, the position of the farmer under rising and falling prices is worth special mention. Rising prices and improved business conditions mean higher prices for farm products, and higher prices for seed, fertilizer, machinery, and the goods farmers consume. But as a matter of observed fact, with improving business conditions and rising price levels, prices of farm products have usually risen more rapidly than either the prices of productive agents used by farmers or the retail prices they have to pay. Thus, the real income of farmers is likely to be markedly increased in periods of rising prices. Moreover, if the farmer has a mortgage his payments are easier since they involve giving up less purchasing power with higher prices. And if the farmer owns his farm, he will be able to sell or rent the farm at a higher figure on account of improved profit prospects in farming. Few classes gain more from rising prices than farmers.

But the picture is reversed if prices fall and business conditions worsen. Just as farm product prices usually rise very rapidly in boom periods, so they fall more rapidly than most other prices with worsening conditions. Since farm product prices fall more rapidly than the prices of productive services used by farmers and than retail prices paid by farmers, the real incomes of farmers are drastically decreased in falling price periods. Moreover, many farmers have mortgages, and falling prices make these mortgages harder to meet, especially when the mortgage has been made in a period of high farm prices and land values. Even though the farmer owns his own farm, he is unable to sell or rent it except at sharply decreased figures. Owing to the very flexible nature of farm product prices relative to prices farmers have to pay, farmers are likely to be more drastically affected than most other groups by booms and depressions, by rapidly changing price levels.

This sketchy summary of the effects of changing price levels on various

groups should indicate especially two things: first, that changing price levels affect different groups differently only because different prices and incomes change at the different rates; and second, that the whole problem is closely tied up with the problem of booms and depressions, a relation that remains to be explored in more detail later.

CHAPTER 41

Spending, Trade, and the Price Level

The Equation of Exchange

IN DISCUSSING the interrelations of monetary conditions, price level changes, and general business conditions, it is helpful to start with a very simple equation that provides a tool for investigating the major factors at work. This equation is first set forth briefly, indicating the interrelations involved, and then each of the factors included is considered at more length in following sections.

Statement of the equation

The "equation of exchange" in its simplest form is usually written $MV = PT$. M stands for the amount of money and credit; V stands for the velocity of circulation of this money and credit; P stands for the price level; T stands for the number of units sold, sometimes called "transactions." The whole equation applies to some definite time period under consideration.

M is the total amount of money and credit in existence during the period under consideration. But this quantity, M , may of course change hands frequently or only slowly. Suppose you spend a dollar in the drugstore this morning. It is later given as change to another customer, who in turn spends it for groceries. Later the groceryman pays it to a clerk as part of his week's wages, and the clerk pays it to the clothing merchant for a new shirt. The one dollar which you had this morning has changed hands several times, carrying on four dollars' worth of transactions (three purchases and one wage payment). In this case the "velocity of circulation" of the money was 4 for the time period considered, and it is this velocity for which the V in the equation stands. To find the total volume of expenditures in any given time period it is necessary to multiply the amount of money and credit by its average velocity of circulation. This MV is sometimes called "total monetary demand"; it is the total amount of expenditures in the period.

During the same time period, the total number of units of goods and services sold may be represented as T . The total number of units sold, multiplied by the average price per unit, gives the total receipts of sellers of goods and services. The price level is simply an average of all prices;

therefore P stands for the average price per unit, which is simply another way of representing the price level, and PT represents the total amount received by sellers in the given time period. From this it is clear why the "equation" necessarily holds true, for it simply states that the amount spent in a given time period is the same as the amount received. It should be evident that the equation is merely a truism—it tells nothing whatsoever about which are the causal factors and which the results. It is useful only because it facilitates consideration of the interrelations of the various factors involved.

If people want to emphasize the price level, they sometimes write the equation, $\frac{MV}{T} = P$, which sets the P alone. Obviously, the equation is still true, because it simply states that the total expenditures divided by the number of units sold gives the average price per unit. Again it is simply a truism—it tells nothing about causal relations.

Interrelations of the factors in the equation

A simple arithmetical example, using hypothetical figures, may be used to show some of the possible interrelations of the factors in the equation. Suppose that M is \$5,000,000; that V is 5 per year; and that T is 25,000,000 units sold. Substituting these figures in the equation we should get $5,000,000 \times 5 = P \times 25,000,000$, so that P would be 1 in this case.

Suppose first that the total amount of money and credit remains the same, but that the people spend their incomes faster after receiving them. This means an increase in velocity of circulation (V), because now each dollar, on the average, gets spent more times during the year. This increased total monetary demand will mean either that prices rise or that more units are sold at the same prices, or both to some extent. Since the equation is a truism, the increased monetary demand must be reflected in either increased P or T , and this of course is only common sense. If the number of units sold rises in proportion to the increased monetary demand, there will be no increase in the price level. •

Putting this case in arithmetical terms, suppose V increases to 10. The MV will be \$50,000,000, and PT will rise accordingly. If the T remains constant at 25,000,000, the price level will double, rising to 2 with the doubled monetary demand. If the T should double, there would be no change in the price level. Actually, probably both P and T would rise with such an increase in monetary demand.

In case of increased monetary demand, is the effect likely to be on the price level or on the number of units sold? Very loosely, one may say that if the increased monetary demand comes in a period of full employment of men and machines, the result will be primarily on the price level, since it is very difficult to produce more goods and services for sale. If it comes in periods of widespread unemployment, the result is more likely to be an increase in the number of units produced and sold. MV is the amount offered for *all* goods and services; T is the number of units of *all* goods and services supplied. When the amount offered for a single

good or for all goods and services together increases, either there will be an increased number of units supplied for sale or the price per unit will rise. There is a close parallel between the case of the individual price and that of the price level, when thought of in this fashion.

In this example it was assumed that for some reason V increases, and possible results were indicated. It would have been equally reasonable to begin by assuming a change in M or T , or by assuming decreases in any of these factors. If M increases and V remains constant, then either P or T will increase in the manner indicated above. An increase in the amount of money and credit and an increase in velocity have the same type of effect on trade and the price level. If T were to increase for some reason, without any increase in total monetary demand, the result would be lower prices. The amount offered would have increased while money expenditures remained constant, resulting in falling prices; and conversely with an assumed independent decrease in T . If the equation is written $\frac{MV}{T} = P$, it may loosely be considered a sort of demand and supply analysis for *all* goods and services, with P the resulting average price.

Care must be taken, however, to avoid the assumption that the price level is simply the effect of causal changes in the other factors. A change in monetary demand may lead to an increase in T or P or both, but an increase in T or P may equally well lead to an increase in total monetary demand. To carry on the simple example of the preceding paragraphs, the higher prices and increased volume of sales are likely to mean greater business activity and employment, which means that incomes are larger and optimism more prevalent. V increases because entrepreneurs spend their receipts more rapidly on wages, raw materials, new machinery, and so on, and income-receivers likewise spend more freely on consumption. The increased P and T are thus likely to lead to an increase in V . M is also likely to be increased, for in such periods banks are likely to make new loans and investments, thereby expanding bank credit. This increased monetary demand (MV) in turn leads to either increased sales or higher prices, which in turn are likely further to increase expenditures. It is essential to see the close interrelation of all the factors. Each is likely to be both cause and effect of the others.

Although attention will be centered primarily on short-period business fluctuations, the equation of exchange is useful also in discussing long-period price level trends. For example, from the close of the Civil War to roughly 1900 there was a gradual fluctuating downward movement of the price level. Through this period there was a steady and substantial increase in the quantity of gold as gold was mined, and this increased amount of gold led to a roughly corresponding increase in money and credit in circulation. At the same time there was a gradual increase in the velocity of circulation of money and credit, owing partially to improved communication and transportation, which permitted easier commercial

and financial relations and partially to changing business customs. Thus there was a definite upward movement of MV . But at the same time there was a steadily and rapidly rising volume of units sold, from the new factories opened and new lands put in cultivation by and for the rapidly growing population. This increase in volume of sales (T in the equation of exchange) was more rapid than the increase in MV , so that in spite of the increase in total expenditures the price level gradually declined. The units being offered for sale were increasing more rapidly than the total amount being offered for them.

Following 1900, the sharp increase in physical output continued, but about then there were several important gold discoveries. These gold discoveries led to a more rapid rate of increase in the quantity of money. At the same time the use of bank checks became more and more common, which further increased M . Also the average velocity of circulation continued to increase gradually. The result was that now MV was growing faster than T , so that the price level began to rise gradually instead of falling. Total monetary demand was increasing more rapidly than the units offered for sale. This was roughly the situation until the beginning of the First World War.

Restatement of the equation

The equation of exchange is a somewhat more useful tool of analysis if written in a slightly different form. So far M has been used to represent both money and credit, and V to represent the average velocity of money and credit together. In analyzing monetary changes and business fluctuations, it is advantageous to have money and credit separated, especially in view of the very different factors controlling the supply of money and of bank credit. Therefore, it is convenient to represent money alone by M , and to use M' to represent credit (mainly bank credit). In the same way the velocities of money and credit may be distinguished, representing the velocity of money by V and the velocity of credit by V' .

The equation is then written $MV + M'V' = PT$, or $\frac{MV + M'V'}{T} = P$.

This is exactly the same equation as the simpler form. The only difference is that the new form breaks down total monetary demand into money and credit separately, instead of lumping them together. The same statements about the interrelations of the factors hold for either equation. Both are simply tools of analysis for discussing the relationships of the various factors at work.

The following sections of this chapter briefly analyze the factors underlying or determining each of the elements in the equation of exchange. Actually it is seriously unrealistic to discuss each of the elements in the equation separately, in view of their very close interdependence. The following sections only set forth some of the forces at work determining each factor, but this approach may be useful in arriving at the complex interdependence of the multitude of factors at work.

The Quantity of Money (M)

Determinants of the quantity of money

The major factors determining the quantity of money (M) have varied considerably according to the time and country under discussion. In the Middle Ages the quantity of money was at the discretion of the ruling monarch, subject to the forces of trade with other areas. In many cases some standard commodity has served as money. Early in the history of this country each colony issued its own money. From then until less than a decade ago, individual private banks had the power to issue paper money, called bank notes. During the first half of the last century the numerous banks formed under state charters issued such notes to great excess, resulting in a flood of almost worthless money, especially in the western states. Although the power of state-chartered banks to issue bank notes was eliminated in 1865, the power of nationally-chartered banks to issue such notes was continued and encouraged until the monetary legislation following 1933. Since that time these "national bank notes" have been gradually retired from circulation until at present few remain, and the power of issue of actual money has finally been centralized in the hands of the federal treasury and the Federal Reserve Banks. The private banks still have, however, great power to increase or decrease the quantity of credit (M'), which in volume is much more important than the quantity of actual money (M).

What determines the amount of money in circulation? During the last century in many countries the presumption has been that, under the gold standard,¹ the quantity of money would vary directly with the amount of gold owned by the government. If more gold came in, either it would be used as money directly or it would serve as the basis for more money and credit. This was to an important extent the case in the United States through the last century and to a lesser extent even into the 1920's. The present situation, however, is quite different. Neither the United States nor any other country is any longer on a gold standard whereby the amount of gold directly controls the quantity of money. Gold has been moving into the United States in tremendous quantities in recent years until we now have over three fourths of the world's monetary gold supply. In spite of this great inflow, there was a far less than corresponding increase in the quantity of money, and the currency increase beyond the level of the 1920's has come with increased hoarding and, recently, better business conditions without any apparent connection with the gold inflow.

Today gold is used as a monetary base only very indirectly, and no gold is coined. There are metal coins of silver, nickel, and copper in circulation; but these are of very minor importance in the total money supply. Paper money constitutes the bulk of our money, and the quantity of paper money is connected only loosely with gold. Three

¹The gold standard is discussed in Chapter 42, "Gold and the Price Level."

major kinds of paper money may be issued, with partial gold and silver backing or with no specific metal backing—but in no case can the gold “backing” be obtained in exchange for the paper money, except for use in international transactions or for use in the arts (jewelry, dentistry, and so forth). The two most important types of money now issued are Federal Reserve notes and silver certificates. The former are issued by the Federal Reserve Banks, which are closely controlled by the federal government. These notes have a minimum gold backing of 40 per cent, but the amount of gold actually back of them may be much more than this—even over 100 per cent as is now the case. The amount of Federal Reserve notes outstanding is largely determined by how many notes people want for use and for hoarding. Those in control may change this amount in response to changed demand for currency or for other reasons. Federal Reserve notes outstanding at the middle of 1942 were \$9,310,000,000. Silver certificates issued by the federal treasury amounted at that time to \$1,754,000,000. All together the amount of money then outstanding amounted to \$12,383,000,000, which was triple the \$4,000,000,000 customarily outstanding in the 'twenties. Part of this increase is accounted for by increased hoarding, part by the increased needs for currency and coin in view of sharply improved business conditions, and part by changed payment habits, occasioned to some extent by increasing bank service charges. But in spite of this increase, not only could the issue of Federal Reserve notes, silver certificates, and metal coins legally be greatly expanded, but in addition up to \$3,000,000,000 of new United States notes (“greenbacks”), with no specific metal backing, can be issued by the Treasury at the direction of the President under powers conferred on him as part of the 1933 emergency legislation.

Although determination of the amount of coin and currency in circulation is within the discretionary powers of the Treasury and Federal Reserve authorities, the amount is in fact determined largely by the demands of the public for such money for current use and for hoarding. As will be explained in discussing the Federal Reserve System, a flexible system exists for supplying such coin and currency as the public desires. If more money is issued than is wanted, it largely flows back to the banks and takes the form of added bank reserves. But this statement that the amount of coin and currency in circulation depends now largely upon the demands of the public for it must not be confused with the question of the total volume of money *and credit*, which is in large part a business cycle and overall monetary and credit control problem.

The federal monetary authorities may use their power to issue money as a means of attempting to expand total monetary demand. Although the public generally will not hold more money than it “needs” for hoarding and for business purposes, by issuing more money than this amount, the monetary authorities may build up bank reserves and provide the basis for further credit expansion. But the present authorities have shown little inclination to use these discretionary powers to increase the total volume of money beyond the demands of the public

for it, even during the depression years when an expansion of total expenditures was greatly desired.

The quantity of money and inflation

With our unprecedented gold stock of about \$23,000,000,000 and our silver stock of another \$4,500,000,000, the amount of money in the United States in 1942 could have been doubled and still have had 100 per cent metal backing. Legally, the quantity of money could be increased much more than this, since far less than 100 per cent metal backing is required by law (or by tradition). Under present circumstances, such an increase in the quantity of money would undoubtedly increase the upward pressure on the price level, in view of the very small volume of unused capacity and unemployed workers now available. That the mere fact of any given amount of gold and silver backing does not assure any given price level is now recognized by the monetary authorities, and a cautious policy toward increasing the quantity of money has been followed by them, though the same cannot be said about their attitude toward bank credit expansion since the beginning of the Second World War.

However, this conservatism in the issuing of paper money has not always characterized government policy in the past. Perhaps the most notable price inflation due in considerable part to excessive issuance of money in the United States was during the Civil War. The Union government at that time issued "greenbacks" to meet the expenses of the war—and in such large quantities that price levels rose and the value of money fell sharply. The inflation precipitated by paper money issued in the Confederacy was much more severe, owing to a relatively much larger issue of money. The most extreme case of this kind of inflation in history was in postwar Germany, where the bankrupt government issued so much paper money that it became almost completely worthless. It took a wheelbarrow full of paper money to buy a postage stamp in 1923, and most of the printing presses in the country had been commandeered by the government to use in printing more money.

These cases are instructive as an indication of the results of such extreme policy and should serve as a warning against such excesses, but they do not mean that every increased issue of paper currency will necessarily have the same result. There is no likelihood that anything approaching such wild inflation will take place in the United States from the use of paper money, at least not in the foreseeable future. If it should come, inflation is much more likely to take place through expansion of bank credit and through changes in the velocity of circulation than through an increased quantity of actual money. A much larger volume of bank credit exists so that any given proportional increase there would be more important. There is also more popular fear of "inflation" from increases in the amount of money than from credit expansion, which makes the authorities reluctant to use the right to issue greenbacks, or to expand other note issues.

In connection with past price level rises accompanying sharp increases in the quantity of money, one sometimes hears references to the "quantity theory of money." Although the "quantity theory" has a variety of interpretations,² the meaning the words usually are intended to convey is this: that an increase in the amount of money will lead to a rise in the price level, and a decrease in the amount of money will lead to a decrease in the price level. Sometimes it is said that an increase in the quantity of money will cause a just proportional increase in the price level, and similarly for a decrease. Persons holding this "theory" point to the inflations during the Civil War and in postwar Germany as proof that their theory "works." However, the preceding discussion reveals the fallacy in this oversimplified analysis of the effects of a changing M on the price level. It would be correct to say that an increase in the quantity of money will lead to higher prices *if* all other factors remain unchanged. But it is by no means certain that all other factors will remain unchanged with an increased supply of money. Especially is this true in periods of low productive activity and unemployment, when increased monetary demand is much more likely to result in increased production and sales than in sharply higher prices. Price inflation is likely as a result of increased issue of money when more goods and services are not forthcoming to meet the increased demand; it is not likely when unused productive capacity exists to turn out more goods in response to increased monetary demand. As a matter of fact, the most drastic price level increases have been associated with sharp increases in the quantity of money; but this should not lead us to the oversimplified conclusion that an increasing quantity of money necessarily means a rising price level, much less one rising just in proportion to the increased quantity of money. Obvious as this may seem, this loose reasoning is one of the commonest of all economic fallacies, not only among untrained laymen but also among government officials in high positions.³

The Quantity of Credit (M')

The quantity of credit has already been discussed in considerable detail in the chapter on "The Banks and the Quantity of Credit." Although there are numerous other types of credit that temporarily serve as payment, bank credit was emphasized since bank checks make up the bulk of our means of payment. Furthermore, bank checks differ from most other credit instruments in that they serve as final payment (just as does money) while most other types of credit instruments merely postpone final payment in money or bank check. This is so because we

² Many of the interpretations are much less lopsided and more defensible than the version noted here. Of course the quantity of money is one of the important factors influencing the price level.

³ A relevant example will be noted presently in connection with the "devaluation" of the dollar in 1934.

do so much of our business by checks that for the most part actual money is seldom demanded from the banks by persons receiving checks.

As was pointed out in the earlier chapter, bankers are led by the desire for profit to make new loans and investments whenever they have excess reserves and can find what they consider safe and profitable borrowers. In periods of pessimism and financial stress, the consideration of safety is uppermost, and credit is likely to be contracted sharply as bankers call in loans and liquidate investments. The ultimate determinants of the quantity of bank credit are all the factors influencing both the bankers' decisions on loans and investments and the potential borrowers' decisions as to how much they wish to borrow or repay. But whatever the bankers' and borrowers' decisions, bank credit cannot be expanded unless excess reserves exist, and the shortage of reserves may set the limit to bank credit expansion if banks have expanded loans and investments near the legal reserve limits.

From the previous analysis of bank credit expansion, it should be easy to see that the quantity of credit is likely to increase in periods of improving business conditions. In such periods businessmen and investors find profit expectations good and are anxious to borrow to finance new projects, even though the same projects may have looked unattractive in depression periods. Likewise, in such periods bankers see the profit possibilities of potential borrowers and become much more willing to make new loans and investments. Risks that make loans appear "unsafe" in depression periods appear "safe" and attractive in the light of improving business conditions and expanding monetary demand. Especially are bankers anxious to extend credit if excess reserves are large. Working with the increases in V and V' that come in such periods, the increase in the quantity of bank credit is likely to be cumulative with better business conditions, increased production and employment, and rising prices, until excess reserves are drawn down to the point where their scarcity exercises a check on further expansion.

Similarly, in periods of deepening depression the quantity of bank credit is likely to be sharply reduced as bankers call in loans and attempt to liquidate investments. Not only are the banks likely to have cash withdrawals at such times, which forces them to contract credit because of reserve shortages, but the bankers become fearful of the safety of loans and investments even if reserves are well above the legal limit. Risks that looked most attractive in the light of boom period expectations often look very shaky in the colder light of recession or depression. Indeed, the shift from favorable to unfavorable expectations may well lead to much greater contraction of loans and investments than appears later to have been necessary to protect the banks' positions. Even if the bankers were willing to lend, few are the entrepreneurs or investors wishing to undertake production and new investments requiring increased borrowing in such a period. Just as bank credit expansion is called forth in good times and itself plays a central part in bringing increased production and higher prices, so bank credit contraction is precipitated

in bad times and itself plays a central part in bringing further contraction of production and falling prices. This intimate interdependence of general business conditions and the supply of bank credit is one of the most vital for the explanation of booms and depressions.

Although they are of much less cumulative importance, other types of credit can and do play a somewhat similar role. For example, in good times installment credit is extended freely, with the result that purchases are made that would have been impossible without installment credit. This increased quantity of credit has a general effect on business conditions and prices similar to that of an increase in bank credit. In bad times, when incomes and employment have fallen, installment buyers are often pressed for payment; fewer installment sales are made. In this way, the consumer credit extended during good times is contracted or destroyed, with the same general effect as a contraction of bank credit. The extension of credit by sellers on open book accounts, on promissory notes, or any other type of credit, and the contraction of this credit by demanding payment and refusing to extend new credit, have the same sort of effects. It is extremely important to recognize the cumulative nature of expansions and contractions in the total volume of credit. Once credit expansion gets under way, it helps create business conditions conducive to more credit expansion, especially by the banks but also by sellers of goods. Once credit contraction gets well under way, it helps create business conditions conducive to more credit contraction, especially by the banks but also by sellers who have extended credit.

Although the quantity of money is under close government control and may be governed with a view toward business stabilization, the quantity of credit rests largely on the actions of private bankers and sellers, motivated primarily by the desire for profits. Therefore, since credit is far more important than actual money in the total volume of transactions, the supply of our means of payment is primarily dependent upon the actions of private individuals and businesses in their everyday banking and business operations. And the quantity of credit, based upon the perfectly natural desire of bankers and businessmen for profits and safety, is likely to move just perversely to what is needed to maintain business stability. Although the Treasury and the Federal Reserve authorities now have substantial means of control over bank credit expansion (to be discussed in later chapters) these controls have proved to be of only limited power, especially in checking credit contraction, once deflation is well under way.

The Velocity of Circulation of Money and Credit (V and V')

It is simple enough to see that the velocity of circulation of money may be more than one, but it is perhaps more difficult to see how a bank deposit may have a velocity of more than one. A dollar bill may be passed freely from hand to hand in payment, but typically a bank check

is only spent once. In fact, however, a bank deposit may have any velocity of circulation.

If you receive a check from Mr. X, you may simply endorse it and give it to the grocer in payment for your groceries. When you endorse it, you are adding your credit to that of the signer of the check. If the check should be no good when presented at the bank, the grocer could collect from you. Simply by endorsement, checks can be passed from person to person in payment, thus having a velocity of more than one. But this is of only minor importance. Relatively few checks are used more than once or twice before being presented at the bank for deposit or for payment in cash.

The important manner in which bank credit may have a higher velocity of circulation is somewhat different. Suppose you borrow \$100, taking the \$100 in deposit credit (M'). You buy a horse from Jones, paying him by check, and Jones deposits the check in his bank. Your deposit is down \$100, but Jones' is increased \$100. The velocity of the \$100 checking account is one, and the total amount of bank credit remains \$100 more than before you borrowed the sum. Jones now buys a bedroom suite from the furniture dealer, paying for it by check. The deposit's velocity is now two. This process may be repeated over and over. Since the bulk of transactions are carried on by check, a checking account may have any velocity, depending on the rate of spending of persons who receive it and on how long it stays in existence.

There are two broad factors influencing the rate at which money and credit is spent. The first consists of the business and payment customs of the country and is a factor responsible for gradual long-run changes in velocity rather than for drastic fluctuations. The second is the short-run factor of people's expectations for, and confidence in, future business and financial conditions, and this unstable factor is largely responsible for the sharp changes in velocity during business fluctuations.

The part played by business customs in determining V and V' is easy to see. Suppose all wages were paid only twice a year. Those receiving their incomes in January would hold part of the funds to be spent in May and June. If people are paid weekly, on the other hand, they spend their incomes relatively quickly after receiving them, and their expenditures then become someone else's income. Obviously in the second case the money is used more rapidly; the number of times each dollar is spent during a year is greater. To take another case, suppose it was common to charge all purchases and pay one's bills only once a month. Compare this with a custom of paying for all purchases immediately. Obviously in the latter case the velocity of circulation of money and deposits would be greater. Individuals would hold money for shorter periods before parting with it. Thus the many customs of the business world play an important part in determining the velocity of circulation of money and credit. But such customs change only gradually, and if they were the sole determinants V and V' would not be expected to vary greatly except over long periods.

There occur, however, very sharp fluctuations in V and V' . People and businesses are influenced by a wide variety of considerations in deciding how much of each period's income they spend and how much they hold idle; some of these are relatively stable and others subject to violent change. One major factor, as indicated in the preceding paragraph, is the time elapsing between receipts of income. Others, subject to sharper fluctuations, center around prices, production and employment and around expectations as to the future course of business conditions. Inseparably intermingled are political and social factors that influence the public's attitudes on economic matters. Fluctuations in velocity are part and parcel of general business fluctuations, and can best be fully considered in connection with the discussion of business cycles and the factors determining the levels of production, employment, income, and prices.

If people expect falling prices and decreased business activity it is logical to hold income idle rather than spend it, because with falling prices the value of money increases and with worsening business conditions investments in goods or securities are likely to decline in value. Generally pessimistic expectations, therefore, usually lead to increased amounts of money and deposits temporarily held idle, to take advantage of expected lower prices, to avoid losses on investments, and to provide for unforeseen contingencies during the bad times ahead. It is an observed fact that a large number of persons have more faith in money as a store of value in periods of uncertainty and stress than in securities, goods, or any other investment opportunity for savings. On the other hand, favorable expectations as to prices and business conditions logically lead to the reverse action—to current expenditures before prices rise and before the most profitable investment opportunities are exploited by others. Further, in the light of improving expectations the volume of idle balances required for unforeseen future contingencies usually declines. Thus in bad times, people and businesses, pessimistic and uncertain as to the future, hold increased portions of their incomes idle and even rush to convert goods and securities into money and deposits. In good times the reverse sweep occurs, with increasing portions of incomes being spent promptly and idle funds being put to use to buy goods, services, and securities.

General political and social conditions play an important role in the complex of factors that determine the volume of funds people and businesses hold as temporary stores of value. In periods when even the future of government seems uncertain and social instability is imminent, businessmen are very reluctant to expand operations and individuals and businesses make every effort to protect against the uncertain future. Under these circumstances people, rationally or irrationally, are likely to rush to money as the "safest" asset to hold, especially because of the immediate availability of cash for any use as compared with the greater illiquidity of other assets. Thus political and social insecurity may lead to sharp increases in the amount of funds temporarily held idle,

as during the 1932-33 economic emergency. At such times people often prefer to hold actual cash rather than bank deposits, so that crisis conditions especially may have differential effects on V and V' , a development which may lead to serious banking difficulties because of the different reserves ultimately held behind money and deposits.

When people and businesses hold their incomes longer than usual before spending them, it is commonly said that they are "hoarding." While this term is loosely used in a variety of ways, its simplest meaning is to denote the fact that people hold money and deposits longer on the average before spending them, or, what is the same thing, that velocity of circulation decreases. "Dis-hoarding" describes the opposite situation, where people spend their incomes more rapidly and hold smaller idle balances. It should be understood clearly that hoarding is a decrease in velocity and not a change in the total amount of money and deposits that people hold. All money and deposits are at any time held, or owned, by someone. Simply because people's expectations worsen does not permit them, as a group, to hold larger balances, unless the government or the banks conveniently issue new funds. If the amount of money and deposits remains the same and everyone increases his idle balances, this simply means that funds will circulate more slowly. If everyone were to hold idle his entire balance, economic exchange as we now know it would cease entirely—but there would still be the same amount of money. Conversely, "dis-hoarding" connotes an increase in the rate of circulation of funds, rather than an increase in the total amount of funds in existence.

Such great fluctuations in velocity as have intermittently occurred exercise a direct and important effect on the price level and on the volume of production and employment. A decreased rate of spending forces down prices, production, and employment; an increased rate of spending has the opposite effect. Moreover, sharp changes in velocity often play a perverse role in intensifying business cycles parallel to that of perverse changes in the volume of credit. In bad times, the decline in velocity reinforces the contraction in the volume of production and employment and presses down on prices. These deflationary results in turn lead to greater uncertainty and more unfavorable expectations, further decreasing the rate of spending. In a similar fashion, the cumulative increase in velocity plays a central role in the development of an inflationary boom, once monetary expansion is well under way. Indeed, velocity changes are likely to be more critical than changes in the volume of credit, since cumulative changes in velocity have less definite limits than those in the volume of credit; exhaustion of available bank reserves and liquidation of outstanding "doubtful" credit, respectively, set rough limits to the expansion and contraction of bank credit, but no comparable limits exist for velocity changes.

With this brief summary, it is advantageous to postpone fuller consideration of the interrelations of velocity changes and general business activity to Chapter 44.

The Volume of Transactions (T)

In general it is correct to say that changes in the volume of goods offered for sale reflect changes in monetary demand—that is, that an increase in $MV + M'V'$ can ordinarily be expected to lead to increased offers of goods for sale, and a decrease in $MV + M'V'$ to decreased offers. This is clearly the most important factor in explaining such overall fluctuations in production as occur in business cycles. It is equally true, however, that the greater are production and employment, the more rapidly firms and individuals will spend their funds in buying materials, hiring workers, buying finished goods, and so on. Therefore, upward shifts in T ordinarily lead cumulatively to increases in V and V' . Similarly, increased output generally improves profit expectations, so that an increase in production is likely to call forth increased bank loans, both because businessmen are more anxious to borrow and because bankers are more willing to lend. Prices are also likely to rise in periods of expanding business activity. Whether changes in $MV + M'V'$ and P are primarily causes and changes in T primarily effects, or vice versa, need not concern us here. The factors all interact closely and generally tend to reinforce each other.

While changes in production are the primary factor mirrored by changes in T , it should be noted that T can change significantly without any change in production, especially over short periods. The volume of goods sold is seldom exactly equal to the volume produced in any given time period. At some times part of current production is being accumulated in inventories; at others existing inventories are being drawn on to supplement the sale of currently produced goods. If times are good and prices rising, businessmen are likely to accumulate inventories. If expectations worsen and prices begin to fall, more goods are likely to be offered for sale than are currently produced until inventories are reduced. Over short periods through the business cycle this source of changes in T often represents a vitally important addition to changes deriving from rising or falling levels of production.

There are other ways in which T may increase without any stimulation from the monetary side. Perhaps most important is through improvements in efficiency and new methods of production that make it profitable to increase output even without any increase in monetary demand. Such innovations decrease the cost of production per unit. Increases in output of this sort have been historically very important and promise to continue to be so. For example, the great increase in physical output during the period following the Civil War reflected the widespread introduction of new machine methods, the opening up of vast areas of new land, and rapid population growth. Less important, the merging of industrial processes under one concern eliminates transactions between firms, reducing T independently of monetary changes. For example, if iron ore is mined by one firm, the ore smelted by another, the raw iron converted to steel by another, and the steel made into rails by another, three transfers

(sales) are involved in getting the ore into steel rails. If, instead, one firm had carried on the entire process, from mining to selling the steel rails, obviously the volume of transactions (T) would be less than in the former case. However, such changes in T are usually very gradual; and the sharp changes of T in the business cycle are of a much different nature, closely tied up with monetary changes.

Interdependence of Spending, Trade, and the Price Level

Any factor in the equation of exchange may be causal in motivating changes in one or more of the other factors. Any factor may be changed by changes in any other. In general, any simple causal relation from one factor solely to another (for example, an increased supply of money causing solely a rise in the price level) is very unlikely. Changes in the price level are virtually certain to be both effect and cause of monetary changes and changes in the volume of transactions. Especially is this close interdependence true in cyclical business fluctuations, which are cumulative movements of spending, trade, and the price level, all in the same direction, with the change in each both cause and effect of continued changes in the others. The equation of exchange is primarily useful in providing a guide for discussion. It tells nothing whatsoever about what factor is most important causally.

CHAPTER 42

Gold and the Price Level

Owing to the historical importance of gold in the monetary systems of the world and its possible vital future role, a brief discussion of the relation of gold to monetary conditions is important. At the outset, we should be clear that gold affects the price level *only insofar as it affects one or more of the elements of the equation of exchange*. There is no mysterious or sacred method by which the price level conforms to changes in the gold supply. But gold has often exercised an important effect on the price level through its effect on the amounts of money and credit and their velocities of circulation (the M , M' , V , and V' of the equation). In order to understand the present situation it is necessary first to look backward at the historical position of gold.

The Gold Standard

Over the last century the "gold standard" has been a very important institution in most major countries of the Western world. The common type of the gold standard has been one in which gold itself was not the only kind of money, but in which changes in the total amount of money and credit in existence were roughly proportional to changes in the quantity of gold. As new gold was received by a government, this gold would serve as the basis for a certain amount of new money and/or credit; but the new money and credit might be backed only partially by the new gold. On one dollar's worth of new gold, two, four, ten, or some other number of dollars of new money and credit might be issued by the government or the banks. The important characteristic was that the quantity of money and credit varied roughly proportionately with changes in the gold supply; and, theoretically at least, each unit of money was "worth" (convertible into) a certain number of grains of gold. Until 1934, for example, each United States dollar "contained" 23.22 grains of fine gold.

This, roughly, is the sort of gold standard to which the United States, England, France, Germany, and most other major countries adhered during the last half of the nineteenth century and up to 1914. Following the First World War, during which there was a suspension of the gold standard, almost all countries again adopted a gold standard, but happenings in the last two decades have steadily undermined this standard until

today no important country holds to it. Most countries have completely abandoned the central element of the plan, namely, the necessary correspondence between gold flows and the quantity of money and credit outstanding.

Before inquiring why this abandonment took place, it is useful to examine briefly the functions fulfilled by the gold standard when it was in general use: (1) The primary function fulfilled by the international gold standard was as an organizer of the flow of international trade and finance and as an expediter of this flow through the provision of stable exchange rates.¹ (2) It is often claimed that the gold standard also served as a guarantee against "inflation" for countries adhering to it. This claim, however, is of dubious validity unless very narrowly interpreted.

1. Under the international gold standard, any country "out of balance"—that is, sending abroad more goods, services, and capital than it receives and therefore receiving gold from other countries—will automatically tend to be brought back into "balance" by the inflow of gold received in payment for this excess of exports. Receiving gold, the country will find its monetary supply increased, and, velocity unchanged, there will be a resulting increase in the incomes of its citizens and in its price levels. The countries losing gold, on the other hand, will find themselves with lower incomes and price levels. This combination of circumstances will mean that the latter countries will find it harder to buy from the original out-of-balance country, while the latter will find it easier to buy from the others. This will tend to decrease their purchases from it, and increase its purchases from them, thereby putting them all back "in balance" and checking the gold flow. The flow of gold, through its effect on different prices and on incomes, tended under the gold standard to organize international economic intercourse and keep it on a more or less "balanced" basis. A gold inflow to a country meant higher prices and incomes, making it harder for other countries to buy there and easier for it to buy in other countries. A gold outflow meant exactly the reverse. Whichever way a country was out of balance, the resulting gold flows acted as a more or less automatic corrective, *so long as gold flows were permitted to have their effect on the money and credit supplies and price systems of the countries involved*. When, as later happened, the gold movements were prevented from taking their effect on prices, the automatic process of adjustment broke down. Although this brief description is greatly oversimplified, it shows fundamentally the part played by gold flows in organizing international trade and finance under the gold standard.

The manner in which the international gold standard expedited international trade and finance by providing stable exchange rates cannot be explained yet since foreign exchange has not yet been discussed. How-

¹ For a general examination of the gold standard and international economic intercourse, see the section, "International Adjustments Under the Gold Standard," in Chapter 53. The present section gives only the brief survey necessary to a discussion of the relation of gold and the price level.

ever, this was one of the greatest advantages of the gold standard, since it provided a high degree of certainty for international traders and investors that they could always convert foreign monies into domestic, and vice versa, at a virtually fixed ratio of exchange.

2. The argument that adherence to the gold standard constitutes a guarantee against "inflation" (by which we shall mean, roughly, a rapid and undesirable rise in the price level) is likely to be seriously misleading. It certainly is not true that the gold standard guarantees a stable price level. On the contrary, it *requires* a fluctuating price level if it is to function properly. Each time new gold enters the monetary system from mines or from abroad, it *must* raise the price level if the gold standard is to function; if gold flows out, it *must* depress the price level if the gold standard is to function. If a great deal of gold flows in, the price level must rise until it is so high that it reverses the flow of gold.

As a practical matter, the history of the United States illustrates clearly the fact that the gold standard is no guarantee against sharp price level changes. The graphs in Chapter 40 show clearly the great inflation that took place during and following the First World War, the precipitous drop following that inflation, and the sharp drop from 1929 to 1933, all while our monetary system was directly based on gold.

On the other hand, it is true that adherence to the gold standard would insure against any such runaway inflation as occurred in postwar Germany, if such adherence were politically and economically possible. However, under such drastic circumstances it is virtually impossible to imagine that any country could remain on a gold standard. The drastic monetary disturbances of that period were far more a symptom than a cause of the difficulties of the German economy. The gold standard actually adhered to does have a real advantage in that it takes the management of monetary affairs largely out of the hands of individuals, in whom we may have limited faith, and puts it on a more-or-less automatic basis, according to gold flows. But if we want to remove power from the hands of government authorities lest they unwisely bring on inflation, there may well be better automatic guides to the proper amount of money than the gold supply—for example, outright stabilization of the price level. Modern economists are inclined to emphasize strongly the failings of the gold standard as a guide to monetary policy. Adherence to this standard may have some value in avoiding discretionary control of the price level by monetary authorities, but it does not insure against sharp changes in the price level nor does it permit authorities to take monetary action as needed against cyclical business fluctuations.

The Abandonment of the Gold Standard

It was this desire to "do something" to mitigate business fluctuations that in large part explains the recent abandonment of the gold standard after its postwar readoption. In the midst of postwar readjustments countries were loath to permit their price levels to fall as they lost gold, especially in central Europe where gold was very scarce. This reluctance

is not hard to understand since falling price levels are usually associated with hard times and unemployment. The opposition reached a climax following 1929, first in Europe and then in the United States. With the sharp financial crisis and depression following 1929, large amounts of gold in Europe were withdrawn from circulation for hoarding and for temporary safekeeping in the United States. This severe gold drain forced a tremendous contraction of money and credit, making still more severe the deepening depression. In September, 1931, England, long the financial center of the world, went off the gold standard—she decided that the advantages of remaining “on gold” were no longer great enough to offset her desire to check the deepening depression by expansionary monetary measures. Remaining on gold meant a contraction of the quantity of money as gold was withdrawn; off the gold standard she could expand her money supply to check the depression and falling price level, regardless of gold flows.²

The abandonment of the gold standard by England was a terrific blow to the confidence of the financial world, especially in those troubled times. In the United States, with the price level falling rapidly and depression deepening, unprecedentedly large amounts of gold were withdrawn from the banks, both for hoarding and for shipping abroad to what were hoped to be safer havens. Banks were failing rapidly as the result of this drain on their reserves. Our monetary authorities were forced to choose between staying on the gold standard (contracting money and credit as gold was lost) and expanding the supply of money to check the falling price level and depression. They chose to adhere to the gold standard, and took few real steps to check the internal crisis. Early in 1932 the Glass-Steagall Act was passed by Congress, permitting the issue of money by the Federal Reserve Banks with new types of security backing to supplement the scarce gold backing. This action temporarily eased the tension. But finally, in 1933 as part of the emergency legislation of the Roosevelt administration, all gold was called out of monetary circulation into the Treasury and the United States went officially off the gold standard. Most of the countries of Europe had already taken this step. Some, like France, clung to gold longer, but today no major country retains the prewar gold standard.

The fundamental reason for the abandonment of the gold standard was that countries were no longer willing to let domestic price levels and business conditions take a secondary position to gold drains and inflows as guides to monetary policy. They preferred to give up the gold standard rather than contract further money and credit during depression years. Without the gold standard the quantity of money and credit could be controlled so as to combat most effectively booms and depressions at home. And today there seems little reason to suppose that the nations are likely again to tie their monetary systems rigidly to gold; the advan-

²By abandoning the gold standard she also hoped to gain a special advantage in international trade, to be discussed later.

tages of mitigating domestic booms and depressions bulk much larger than those of the international gold standard.

This does not mean, however, that gold has ceased to be of importance, either domestically or internationally. Even though monetary systems are no longer tied directly to gold flows, gold still performs two major functions: (1) It helps maintain confidence in the money and credit issued; and (2) it serves as the basic means of settling international balances. Even though today no person in this country can get gold for his money (except for international trade or use in the arts), the fact that the government has so much gold is one factor that accounts for confidence being maintained in the money issued by the government, and more indirectly in the banks that provide most of the credit. And every country in the world, save possibly the United States, is anxious to obtain gold in exchange for goods and services since gold still is everywhere acceptable as payment in international trade, even though monetary systems are no longer tied directly to precious metals.

By 1942 the United States government owned roughly \$23,000,000,000 of gold, priced at \$35 per ounce, which was roughly 75 per cent of the total world supply of monetary gold. In 1932, the gold stock of the government was about \$4,000,000,000. The reasons for this huge increase will be discussed in Part XI. Discussion here will be confined to a consideration of the connection between this gold increase, monetary conditions, and the price level in the United States.

"Too Little" and "Too Much" Gold

The problem of "too little" gold

In 1933 the incoming administration recognized clearly the need for raising the price level as an integral part of a recovery program. Partly to eliminate drains on the banks and partly to get all available gold as a basis for new issues of money, it required that all monetary gold be turned in to the government in exchange for other types of money. Gradually also it increased its purchases of newly mined gold, thus stimulating gold production. The purpose of so increasing gold purchases was largely to raise the price level, presumably through increases in the amount of money. Finally, in 1934 the government declared that henceforth a dollar should contain only 13.71 grains fine gold instead of the previous 23.22 grains, a power which had been given the President in earlier emergency legislation. By this action the dollar was "devalued" to 59 per cent of its previous gold value. Since the amount of gold in each dollar was diminished, it now takes more dollars to equal an ounce of gold. At the new dollar gold content, the price of gold is \$35 per ounce, whereas previously it had been \$20.67 per ounce. To say that the price of gold was raised from \$20.67 to \$35 per ounce is identical with saying that the gold content of the dollar was decreased to 59 per cent of its earlier content. This bookkeeping transaction increased the dollar value of the

gold stock of the government from about four billion to about six and a half billion dollars.

Apparently many of the administration's advisers expected that this devaluation of the dollar would immediately result in a rise in the price level of roughly the same proportion. For years it had been commonly declared that the value of the dollar depended on the gold in it—hence decreasing the gold content of the dollar would raise prices correspondingly. Actually the expected result did not materialize. Although the gradual rise in the price level continued, no miraculous upswing ensued. Those who expected it to do so were guilty of a serious oversimplification of the connection between gold and the price level. Assuming that the amount of gold in the dollar determines its purchasing power, they failed to recognize that changes in the gold supply can affect the price level only through their effect on some element of the equation of exchange. The devaluation did not immediately increase the quantity of money (though it did give the government and Federal Reserve more power to do so); therefore it did not raise the price level this way. It had no immediate effect on the quantity of bank credit (though it did increase the power of the Federal Reserve to expedite bank credit expansion); therefore it did not immediately raise the price level in this way. The one way in which it did immediately exert an upward influence on the price level was through its psychological effect. Large numbers of persons, like the advisers advancing the plan, expected the immediate result to be higher prices. Therefore these people rushed to spend their money to make purchases before prices rose, and this increased spending itself helped bring about the expected higher prices. Although the net result of the devaluation probably has been a higher price level than would otherwise have existed over a period of years, the experience is of value in illustrating the dangers of oversimplified cause-and-effect reasoning in regard to gold, money, and prices.

The problem of "too much" gold

More recently "the" gold problem of the United States has been too much gold. It is helpful to trace briefly the course of new gold entering the United States from abroad, in order to see how much effect it may be expected to have on the price level. Under the present law, all monetary gold received by persons or firms here must immediately be sold to the government at \$35 an ounce. Suppose an importer receives \$100 of actual gold. He will take the gold to his bank, receiving credit for a \$100 deposit. The bank then sends the gold to the Federal Reserve Bank with which it deals, increasing its reserves held at the Federal Reserve Bank by \$100. The Federal Reserve Bank then turns over the gold to the federal treasury, receiving in exchange gold certificates (pieces of paper showing that the Federal Reserve has turned over to the Treasury the stated amount of gold). The Treasury then has the gold, as required by law. What may be the effects of this new gold?

There is no particular reason to suppose that the Treasury will expand the amount of actual money it issues just because of this new gold. The

decision as to how much money shall be issued is largely independent of gold flows under the present system. And there is no particular reason to suppose that the Federal Reserve Bank will issue more Federal Reserve notes just because it has now \$100 more gold certificates to serve as backing for new notes. Here again, the decision as to how many notes shall be issued is largely independent of gold flows, although minimum gold requirements are set by law for Federal Reserve notes. Under the present system the quantity of money does not depend on gold flows.

But the picture regarding the supply of bank credit is different. The \$100 gold import increases the importer's bank deposit by \$100; it also increases the bank's reserves by \$100. Since the bank needs to hold only a fractional reserve against its deposits, its excess reserves are increased. If it is a New York bank with a 20 per cent reserve requirement, it now has \$80 of additional excess reserves (that is, the \$100 deposit less \$20 required reserve on it). Thus the inflow of the new gold has had a twofold effect on bank credit. First, it increases the amount of bank credit by \$100, since the importer obtains a new deposit of that amount. Second, by increasing excess reserves it lays the basis for an ensuing multiple expansion of bank credit. This further increase will come if the banks make new loans and investments on their excess reserves; it will not come if the banks simply permit the excess reserves to lie idle.

Thus, even though the United States is no longer on the gold standard, the possible effects of gold flows on our money and credit system, and through them on the price level, should not be underemphasized. The great gold inflow of recent years has correspondingly increased bank deposits (M') as the imported gold is turned over to the government through the banks. More important, it has created a vast mass of excess reserves that have provided the basis for a multiple expansion of bank loans and investments as well as for a rapid increase in money in circulation. In depression periods new excess reserves may well lie unused; in such expansionary periods as that from 1940 to the present they are likely to underlie multiple credit expansion.

In the same way, an outflow of gold would decrease both bank reserves and deposits, wiping out excess reserves. If the banks had excess reserves such a gold outflow need cause no contraction of bank credit; but if loans had been made on these gold reserves, loss of the reserves would necessitate contraction of all the credit extended on them, which might be several times the actual reserves. Actually this latter contingency is unlikely, for the government has the power to restrict gold exports if it desires. Although the government will now sell gold back to people at \$35 an ounce for export, it could cease to do so at any time, and it surely would if the loss of gold promised to disrupt seriously the domestic economy. Alternatively, reserve requirements might be lowered by the Federal Reserve authorities.

Gold, Silver, and the "Mixed Standard"

If the United States is not on the gold standard, what is the basis of our monetary system? At present, legally we have a "mixed standard."

A law passed in 1934 prescribes that the metal backing held by the Treasury for the money it issues shall consist of 75 per cent gold and 25 per cent silver, though the law does not prescribe what the total metal backing for money shall be—it only sets certain minimum metal backings for various types. Since in 1934 the Treasury held less than the prescribed 25 per cent of silver, the government immediately set about buying large amounts of silver and has continued to do so, thus maintaining the price of silver far above what it would otherwise be. Between 1934 and 1942, the government bought over three billion dollars worth of silver.

Actually silver long ago passed from importance as the major monetary metal. Why then does the government continue to buy all this silver, thereby subsidizing silver producers here and abroad by huge sums annually and calling forth greatly increased silver production? The answer is primarily that the silver-producing states have long had an exceptionally able group of senators, and these senators by skillful log-rolling and vote-trading with their fellow senators have put through the legislation requiring the purchase of these large amounts of silver, far above the open-market price. In 1942 the government owned over four billion dollars worth of silver, valued at \$1.29 per ounce.

In spite of the large purchases of silver made in recent years, gold holdings have increased still faster, so that our silver purchases in turn must be increased more if the prescribed ratio between gold and silver is to be achieved. Economically there is very little justification for these vast silver purchases; they must be explained, and justified if they can be justified, as special subsidies to silver-producing interests. The history of American silver legislation during the last decade is an amazing example of the success of a small pressure group in obtaining large scale subsidies regardless of the complete lack of basic monetary justification for the laws enacted.

The relation of metal reserves to the quantity of money and to the price level, then, is mostly indirect: It may influence the decisions of the Treasury and the Federal Reserve Banks to issue or withdraw money; it may influence the amount of credit created by the banks on the reserves that they may have available; and it may influence the decisions of people to spend or hoard their money (depending on their confidence in the money and in business conditions) and their decisions to borrow or not to borrow from the banks. There is a further important relation through the channels of international trade, but discussion of this problem must be postponed until later.

Future Policy in Regard to Gold

Why has the United States recently received so much gold? Are we likely to get "stuck" with the gold? How will this affect our international trade? Such questions as these cannot be answered until the gold problem has been considered in connection with the whole international situation. One specific question can be answered here, however:

Why not immediately return to the gold standard and use this gold as money to pay off government debt or finance the war?

One reason that the United States and other nations hesitate to take this step is the great disadvantage of again tying monetary policies to gold flows. Return to a gold standard would mean relinquishing the power to increase or decrease the quantity of money and credit as desired to eliminate high booms and deep depressions, rather than simply letting the price level and business activity go up or down as gold flows dictate. It is quite possible that after the war gold flows will again be sharp and unpredictable. A monetary system rigidly geared to these gold flows would be likely to lead to disastrous domestic fluctuations. In a peaceful and settled world, the case for a return to the gold standard would be much stronger.

To use the country's gold supply directly to pay for war production would be the equivalent of printing a corresponding amount of new paper money so far as the government spending's upward pressure on prices is concerned. The increased incomes received by those receiving the gold would go to bid up prices of scarce products, just as they would if people had received paper money, coin, or checks. With the danger of inflation uppermost, only by financing government spending by means that reduce private buying power (such as taxes) can a monetary barrier be imposed. Financing the war program by paying out gold would merely increase incomes and the amount of money in circulation, thus contributing to inflationary pressure.

Moreover, to return gold to circulation as money would yield highly undesirable long-term internal results. One of the primary difficulties faced by the banking system during the depression crisis of the 1930's, and during previous financial crises, was the mass demand of the public for gold to hoard. These mass withdrawals of gold during panics not only forced the banks to close in many cases (bringing simultaneously a mass contraction of credit by the banks in order to get reserves and cash to meet withdrawals), but also seriously depleted the gold supply of the federal government. With gold completely withdrawn from circulation many of the undesirable results of gold hoarding are avoided. The monetary policy of the government is not so dependent on the panic-stricken actions of the public. Moreover, no serious loss is incurred by having the gold withdrawn from actual circulation. Paper money serves equally well, and indeed is superior in being much more convenient to handle. So long as paper money is commonly accepted as means of payment, no one needs gold as money.

These last paragraphs attack only one narrow aspect of the gold problem. A more comprehensive discussion can be undertaken after analyzing business cycle and international trade and finance. The present discussion is intended primarily to show certain chief relations of gold, money, and prices and to provide a brief case study in monetary dynamics.

CHAPTER 43

The Federal Reserve System

THE Federal Reserve System was founded in 1914 in order to perform two primary functions. One was to provide certain regular services for the banks of the country, the general public, and the federal government. The other, and the one which is of primary interest here, was to maintain "monetary and credit conditions favorable to sound business activity in all fields—agricultural, industrial, commercial."¹ It is the purpose of this chapter to describe briefly the organization of the Federal Reserve System and to discuss the manner in which it carries out its various duties, with special reference to its powers to maintain monetary and credit conditions favorable to sound business activity.

Organization of the Federal Reserve System

As it is now constituted, the Federal Reserve System is made up of the following:

1. The Board of Governors
2. The twelve Federal Reserve Banks
3. The Federal Open Market Committee
4. The Federal Advisory Council
5. The member banks

Each of these is described briefly in the following paragraphs.

1. *The Board of Governors* is composed of seven members, appointed by the President and confirmed by the Senate. Members receive a salary of \$15,000, and are appointed for 14 years, with one term expiring every two years, in an effort to safeguard the Board as far as possible from political pressure groups. In most matters the ultimate responsibility for major policies of the twelve Federal Reserve Banks lies with the Board of Governors; and since the twelve Federal Reserve Banks in turn supervise and regulate the member banks, ultimate responsibility for the entire system is largely centralized in the Board of Governors. As the

¹ *The Federal Reserve System: Its Purposes and Functions*, Board of Governors of the Federal Reserve System (Washington, 1939), p. 23. This little hundred-page booklet is an excellent reference work for simpler points in regard to Federal Reserve organization and activities. This chapter is based directly upon the booklet in several respects.

system was originally constituted there was much less centralization of authority, since at the time of formation of the system there was great fear of control by the "Wall Street bankers"; but this division of authority between the Board and the twelve Federal Reserve Banks resulted in so much dispute and delay at crucial moments of financial need that in 1935 control was largely centralized in the Board.

2. *The twelve Federal Reserve Banks* each serves a certain district in the United States. The banks are located in Boston, New York, Philadelphia, Cleveland, Richmond, Atlanta, Chicago, St. Louis, Minneapolis, Kansas City, Dallas, and San Francisco, and where the district served is very large some have branches to serve member banks more conveniently. Each Federal Reserve Bank was founded by the sale of stock to member banks, which are required to buy stock, and pays regular dividends of 6 per cent to member banks holding stock. As stockholders, member banks have the power to elect six of the nine directors of each Federal Reserve Bank. Of these six, three may be bankers, and three must represent business, industry, and agriculture within the district. The other three directors are appointed by the Board of Governors. But the policies of the twelve Federal Reserve Banks are largely determined by the Board of Governors, although each Federal Reserve Bank does give some representation of the views and interests of its particular region. And the Federal Reserve Banks are operated in the public interest, not for profit—though so far they have derived from operations income sufficient to pay 6 per cent dividends on all stock, accumulate large surpluses, and pay substantial amounts into the United States treasury, which until 1933 received any such excess profits. During the late 1930's public policy dictated that the Federal Reserve Banks operate so as to make a loss on some transactions, and the Board of Governors has so operated when this action was called for to promote the public welfare.

3. *The Federal Open Market Committee* consists of the seven members of the Board of Governors, plus five other members chosen by the twelve Federal Reserve Banks. This twelve-member open market committee determines the policy of the twelve Federal Reserve Banks in regard to open market operations—that is, the purchase and sale of government securities and other obligations in the open market. These open market operations, which will be discussed presently, are one of the primary means by which the Federal Reserve authorities attempt to control bank credit expansion. Although the Board of Governors does not determine open market policy independently, its seven members (when there are no vacancies) constitute a majority of the open market committee.

4. *The Federal Advisory Council* consists of twelve members, one selected annually by each of the twelve Federal Reserve Banks. This Council meets with the Board of Governors at least four times a year and makes recommendations for Federal Reserve policy. But, as the name suggests, the Council's powers are purely advisory—it has no power to force the Board of Governors to accept any of its recommendations. The major purpose of this Council at present is to make sure the interests of the

twelve Banks and their districts are brought to the attention of the Board of Governors at regular intervals.

5. *The member banks* include all national banks in the United States and those state banks that desire membership and conform to the requirements set up for member banks. At present slightly less than half the 15,000 banks in the United States are member banks, but the 8,000 non-member banks are almost all small banks, representing only about 17 per cent of the total deposits of the banking system. It is likely that some day pressure will be brought on all banks to become members of the system in order that control of the entire banking system may be more closely centralized in federal hands. At present the Federal Reserve has direct supervision over far the largest per cent of banking assets and deposits; but the nonmember banks, controlled by unstandardized state regulations, still constitute a substantial bloc.

As already stated, each member bank is required to buy stock in its Federal Reserve Bank equal to 3 per cent of the member bank's capital and surplus. Each member bank is also required to maintain its legal reserves with the Federal Reserve Bank in its district. In addition, member banks are subject to regular examination by Federal Reserve authorities, and are required to meet the operating standards set by the Federal Reserve.

This brief description of the organization of the Federal Reserve System provides a background for the discussion of the manner in which it carries out its various duties in serving the public, the member banks, and the government, and in maintaining monetary and credit conditions favorable to sound business activity. The regular service functions of the system include primarily the following: holding member bank reserve balances; furnishing money for circulation; clearing and collecting checks; supervising member banks; and acting as a fiscal agent for the federal government. These regular service functions engage by far the largest part of the time and energy of the officers and employees of the Federal Reserve Banks, but they are largely routine duties. The policy decisions involved from time to time in deciding what action should be taken to maintain monetary and credit conditions favorable to sound business conditions are of a very different sort. Such policy decisions are made by the Board of Governors and the Open Market Committee, and are likely to have much more far-reaching consequences than the service functions. They involve primarily open market operations, fixing legal reserve requirements, lending to member banks, establishing rediscount rates, and otherwise influencing member banks and the general public in attempts to maintain or achieve satisfactory business conditions.

Service Functions of the Federal Reserve

Holding member bank reserves

Each member bank is required by law to keep its required reserves with the Federal Reserve Bank in its district. Member banks may of course

also keep excess reserves with their Federal Reserve Banks, and typically they do. These reserve balances kept by member banks with the Federal Reserve Bank are essentially checking accounts of the member banks with the Federal Reserve, just as an individual has a checking account with a member bank. Except that it must always have the legal reserve required behind deposits, a member bank is free to draw on, or add to, its reserve account as it wishes. It receives no interest on its reserves.

The Federal Reserve Bank in turn is required by law to hold gold certificates or money that equals at least 35 per cent of the deposits it holds. Thus, ultimately a gold reserve is required behind all bank credit, but this gold need only be 35 per cent of member bank reserves, which reserves in turn may serve as backing for several dollars of bank credit for each dollar of reserves. Legally one dollar of gold in the Federal Reserve can serve as backing for \$2.85 of member bank reserves held by the Federal Reserve. Assuming a 20 per cent member bank legal reserve requirement as a rough approximation, this \$2.85 of reserves could theoretically serve as the basis for \$14.25 of bank credit extended by member banks. Actually member banks may not expand credit to this extreme legal limit for reasons noted in previous chapters; but it should be apparent that the gold requirement for Federal Reserve Banks against member bank reserves is so small as ordinarily to exercise little restraint on bank credit expansion. There may be occasions, however, when this gold requirement does limit the deposits (reserves) a Federal Reserve Bank can hold, and hence does put a limit to credit expansion. Such an instance occurred in 1931.

Furnishing money for circulation

All actual money in the United States is now created either by the federal treasury or by the Federal Reserve Banks. Treasury currency, consisting of silver certificates, metal coins, and United States notes, is issued by the Treasury but is placed in circulation largely through the Federal Reserve Banks as a means of getting it to the commercial banks and thus into general circulation. The Federal Reserve Banks themselves issue Federal Reserve notes, by far the most important single type of money today. These Federal Reserve notes are liabilities of the issuing Federal Reserve Bank, and also of the federal government. Each Federal Reserve note must be fully backed by collateral held by the issuing bank. At least 40 per cent of this collateral must be gold certificates; the other 60 per cent may be gold, rediscounted paper, government bonds, or other designated acceptable security.

The total amount of paper money and coin in circulation has increased steadily in recent years, owing largely to increased hoarding and to increased business activity. But aside from this special increase, the furnishing of money for circulation is a regular operation. New currency is constantly being put into circulation to replace old, worn currency. There are reasonably predictable increases in the amount of currency in circulation at certain periods of the year, such as Christmas, harvest

time, around the Fourth of July, and so on, when people and businesses want more hand-to-hand cash. This cash is put into circulation simply by the banks paying it out on withdrawals made by customers during such periods. The member banks get the cash to pass out to customers by drawing on their reserve accounts at the Federal Reserve Banks. The Reserve Banks always keep on hand large supplies of all sorts of paper money and coin to meet needs of member banks. Following these periods of increased demand for cash, the cash flows back to the member banks and is returned by them to their reserve accounts in the Reserve Banks. Thus the Federal Reserve stands ready to furnish cash for circulation whenever needed by the member banks. And even if member banks do not have excess reserves on which to draw for cash, the Reserve Banks may increase member banks' reserves by loaning them money or buying securities from them and in this way make cash available to the member banks. These ways of increasing member bank reserves are discussed under the powers of the Federal Reserve authorities to control money and credit conditions.

Although the total quantity of money has ordinarily fluctuated relatively little around its steady increase, the Board of Governors, acting through the Federal Reserve Banks, has the power to increase Federal Reserve notes up to the legal limit set by the 40 per cent gold requirement and the 100 per cent total collateral requirement. Ordinarily these legal requirements leave considerable leeway for increase of the notes outstanding, so that ordinarily the issue of Federal Reserve notes is largely a matter of discretion of the Federal Reserve authorities. As stated above, they commonly stand ready to supply the currency needed for hand-to-hand circulation. But this discretionary control of the amount of currency also gives them a means of influencing the price level and general business conditions. This power of increasing or decreasing Federal Reserve notes in circulation as a price level control device has been sparingly used. The Board of Governors has much more commonly attempted to influence the price level and business conditions by controlling the amount of bank credit, largely because bank credit is quantitatively so much more important and because money issued beyond the public's "wants" will simply be returned to the banks.

Clearing and collection of checks

Almost all bank checks drawn on out-of-town banks are cleared through the Federal Reserve System. For example, suppose a merchant in Chicago sells a \$100 bill of goods to a customer in Detroit, and receives in payment a check on the customer's Detroit bank. The merchant, instead of having to go to Detroit to collect the check, simply deposits it at his bank in Chicago, receiving a \$100 credit to his account. The Chicago bank then sends the check to the Chicago Federal Reserve Bank, which credits the reserves of the Chicago bank by \$100 and debits the reserves of the Detroit bank by \$100. The check is then sent to the Detroit bank, which debits the customer's account by \$100. The Chicago

merchant has his \$100; the Detroit customer has \$100 less. And since both banks keep their reserves with the Chicago Federal Reserve Bank, the process of clearing the check is accomplished simply by increasing the reserve account of the Chicago member bank and decreasing the reserve account of the Detroit member bank.

If the check had been written by a customer in New York, instead of Detroit, the collection process would be similar although it would involve two Federal Reserve Banks instead of one, because New York and Chicago are in different Federal Reserve districts. The Chicago merchant would again take the check to his bank and receive deposit credit of \$100 for it. The Chicago bank would then send the check to the Chicago Federal Reserve Bank, which would increase the Chicago member bank's reserves by \$100. Then the Chicago Federal Reserve Bank would send the check to the New York Federal Reserve Bank, where the account of the New York bank upon which the check was drawn would be debited by \$100. The check would then be sent to the New York member bank, which would debit the account of the customer drawing the check. Again the Chicago merchant has received his \$100, and the New York buyer has \$100 less.

This process would complete collection of the check, but it would leave the New York Federal Reserve Bank owing \$100 to the Chicago Federal Reserve Bank. To take care of such cases, the twelve Federal Reserve Banks maintain in Washington an "Interdistrict Settlement Fund." This fund is composed of accounts kept there by each of the twelve banks. Then when the New York Bank owes the Chicago Bank \$100, this sum is simply transferred from the account of the New York Bank to that of the Chicago Bank within the Interdistrict Settlement Fund without any shipment of currency. Since checks payable from one district to another ordinarily about offset each other, these interdistrict clearings transfers are ordinarily fairly small in total. Such transfers as are necessary to settle adverse clearings balances between the Federal Reserve Banks are made by telegraph through the Interdistrict Settlement Fund.

Nonmember banks may use the Federal Reserve clearing system, and many of them do. Nonmember banks using the clearing system are required to maintain "clearing balances" with the district Federal Reserve Bank in order to give the Federal Reserve Bank a fund on which to draw to pay checks drawn on the bank and to which to add checks payable to the bank. To give some idea of the magnitude of this clearing function, in 1938 the Federal Reserve System handled roughly a billion checks, the total value of which was \$232,000,000,000.

Supervising member banks

Banks in this country are subject to supervision by several authorities. The Federal Reserve supervises all member banks. The Comptroller of the Currency supervises all national banks. Each state supervises all state banks chartered by it. The Federal Deposit Insurance Corporation

time, around the Fourth of July, and so on, when people and businesses want more hand-to-hand cash. This cash is put into circulation simply by the banks paying it out on withdrawals made by customers during such periods. The member banks get the cash to pass out to customers by drawing on their reserve accounts at the Federal Reserve Banks. The Reserve Banks always keep on hand large supplies of all sorts of paper money and coin to meet needs of member banks. Following these periods of increased demand for cash, the cash flows back to the member banks and is returned by them to their reserve accounts in the Reserve Banks. Thus the Federal Reserve stands ready to furnish cash for circulation whenever needed by the member banks. And even if member banks do not have excess reserves on which to draw for cash, the Reserve Banks may increase member banks' reserves by loaning them money or buying securities from them and in this way make cash available to the member banks. These ways of increasing member bank reserves are discussed under the powers of the Federal Reserve authorities to control money and credit conditions.

Although the total quantity of money has ordinarily fluctuated relatively little around its steady increase, the Board of Governors, acting through the Federal Reserve Banks, has the power to increase Federal Reserve notes up to the legal limit set by the 40 per cent gold requirement and the 100 per cent total collateral requirement. Ordinarily these legal requirements leave considerable leeway for increase of the notes outstanding, so that ordinarily the issue of Federal Reserve notes is largely a matter of discretion of the Federal Reserve authorities. As stated above, they commonly stand ready to supply the currency needed for hand-to-hand circulation. But this discretionary control of the amount of currency also gives them a means of influencing the price level and general business conditions. This power of increasing or decreasing Federal Reserve notes in circulation as a price level control device has been sparingly used. The Board of Governors has much more commonly attempted to influence the price level and business conditions by controlling the amount of bank credit, largely because bank credit is quantitatively so much more important and because money issued beyond the public's "wants" will simply be returned to the banks.

Clearing and collection of checks

Almost all bank checks drawn on out-of-town banks are cleared through the Federal Reserve System. For example, suppose a merchant in Chicago sells a \$100 bill of goods to a customer in Detroit, and receives in payment a check on the customer's Detroit bank. The merchant, instead of having to go to Detroit to collect the check, simply deposits it at his bank in Chicago, receiving a \$100 credit to his account. The Chicago bank then sends the check to the Chicago Federal Reserve Bank, which credits the reserves of the Chicago bank by \$100 and debits the reserves of the Detroit bank by \$100. The check is then sent to the Detroit bank, which debits the customer's account by \$100. The Chicago

chasers, and crediting these payments to the government's account. When securities are paid off, the Reserve Banks again take over the task of making the actual disbursements, debiting the government's account. The New York Federal Reserve Bank acts as agent for the Treasury in its foreign exchange dealings. Some 2,500 of the system's 11,000 employees devote their full time to carrying out these fiscal functions for the government. The Treasury and other federal agencies served by the Federal Reserve Banks reimburse the Banks for much of the expense involved in carrying out these services.

Credit Control by the Federal Reserve

Although the service functions just listed occupy the time of the vast majority of the Federal Reserve's personnel, the function of maintaining "monetary and credit conditions favorable to sound business activity in all fields" is likely to have much more far-reaching effects on economic life. Primarily the Federal Reserve authorities attempt to carry out this function through influencing the amount of credit extended by the banking system. Aside from the power of currency issue, there are seven major means available to the System authorities for controlling the amount of credit extended.

Open market operations

The primary means of controlling bank credit used by the Federal Reserve over the last decade has been the purchase and sale of government securities in the open market of the New York stock exchange—hence the name "open market operations." By buying government securities, the Federal Reserve can increase banks' reserves, which is conducive to better business conditions and higher prices. By selling government securities, the Federal Reserve can draw down banks' reserves, which tends to check expanding credit and rising prices.

Suppose first the Open Market Committee feels that an inflationary boom is in progress and wishes to impose a restraint on bank credit expansion in order indirectly to check the too-rapid boom and rising prices. It will then sell securities in the open market. These securities may be bought by individuals, businesses, or the banks themselves. Suppose first an individual or nonbanking institution buys a \$1,000 bond sold by the Federal Reserve. The buyer might pay the Federal Reserve in currency, though this would be very unusual. If he did pay in currency this would withdraw \$1,000 of currency from circulation, which would help check the boom.

Far more likely, the buyer would pay for the bond with a check on his bank. The Federal Reserve would then collect the check from the member bank on which it was written, decreasing the bank's reserves by \$1,000. The member bank's deposits would be down \$1,000, and its reserves down \$1,000. It thus loses not only the reserve behind the lost \$1,000 deposit (say 20 per cent or \$200 for city banks), but also the additional \$800 that had been either excess reserves or reserve behind

other deposits. When the buyer pays the Federal Reserve for the bond, this decreases the excess reserves of the buyer's bank by a sum almost as large as the purchase—here by \$800 on the \$1,000 bond purchase. Such a loss of reserves of course tends to check bank credit expansion. If the banks themselves buy the bonds sold by the Federal Reserve, the restrictive effect on credit expansion is stronger yet. Suppose a bank buys the \$1,000 bond. In paying for its purchase it loses \$1,000 of reserves, but its deposits remain unchanged. Therefore it has lost the full \$1,000 of excess reserves, an even greater loss than when a customer of the bank bought the bond. By decreasing the excess reserves of member banks, the Reserve authorities can decrease the credit expansion possible and put direct pressure on member banks to go more slowly in making new loans and investments.

Sale of securities by the Federal Reserve is likely to be effective in limiting bank credit expansion if it draws excess reserves down to low levels. Hence it is likely to be most effective in times when banks have already made large loans and investments. It will be much less effective if banks have large excess reserves on hand, for they can then lose reserves and still have left an adequate base for new loans and investments. The power of the Reserve authorities to check credit expansion by open market operations thus ultimately depends on the quantity of securities held by the Reserve Banks relative to the excess reserves of the commercial banks. For example, in early 1941 excess reserves totaled over \$6,000,000,000, an unprecedentedly high figure, while the Federal Reserve owned only about \$2,500,000,000 of government securities. Thus, the Federal Reserve had only very limited open market power to check a wave of bank credit expansion as banks began to expand credit. The securities of the Federal Reserve would have been exhausted long before excess reserves of the member banks had even run low.

Suppose, on the other hand, that the Open Market Committee wishes to stimulate bank credit expansion in order to stimulate recovery from a depression or check a business recession. It can then order the Federal Reserve Banks to buy securities in the open market. These securities may be purchased either from banks or from individuals and nonbanking institutions. Suppose first an individual sells a \$1,000 bond to the Federal Reserve. He receives in payment a check from the Federal Reserve Bank buying the bond, which check he may either cash or deposit at his bank. If he cashes the check, his bank will obtain the \$1,000 from the Federal Reserve by collecting the check, and the net result will be \$1,000 more currency in circulation, a result which tends to stimulate increased buying and higher prices. If the individual simply deposits the check at his bank, both the bank's deposits and its reserves are increased by \$1,000. Since it needs hold only a fractional reserve against deposits (20 per cent or \$200 to take the previous example) the net result is an increase of \$800 in excess reserves. If a bank itself sold the bond bought by the Federal Reserve, excess reserves would be even further increased, since the bank would receive \$1,000 of new reserves

from the Federal Reserve but its deposits would be unchanged. Therefore, the net result would be an increase of \$1,000 in the bank's excess reserves.

By increasing banks' excess reserves, the Federal Reserve increases the incentive to make new loans and investments, for no banker likes to lose the interest he might earn on idle reserves. However, if excess reserves are already large, open market purchases are likely to be ineffective in stimulating credit expansion. If a bank already has large excess reserves and is failing to make loans and investments because of inability to find "safe" and "profitable" borrowers and because of a general lack of desire to borrow for business purposes at such times, the increased excess reserves resulting from Federal Reserve open market purchases are not likely to lead the bank to expand credit very much, if at all. Like most other Reserve controls, open market policy is likely to be much more effective in checking credit expansion in boom times than in stimulating credit expansion in bad times, although it can exercise some pressure toward expansion by increasing excess reserves that bankers dislike to keep idle.

With the tremendously increased volume of government securities issued to finance depression spending and the present war, it has become an increasingly important aim of Federal Reserve open market policy to help stabilize the bond market to aid Treasury financing. Recently this aim has conflicted with, and largely eliminated, any possible open market operations to curtail credit expansion.

Rediscount rate changes

When a member bank runs short of reserves (because of making many new loans and investments, because of cash withdrawals, because of adverse clearing balances, and so on), it may borrow from the Federal Reserve Bank to replenish its depleted reserves. Ordinarily in such a case the member bank would "rediscount" notes. The member bank has made loans to customers for short periods on the customers' promises to repay, which promises to repay are called "notes," or "commercial paper." Since it deducts its interest in advance, we say the bank discounts its customers' notes. When the member bank needs reserves, it can "rediscount" these notes with its Federal Reserve Bank—that is, it can borrow from the Federal Reserve Bank on the security of such a note, just as the customer borrowed from it on the security of the same note. And the rate of interest, or discount, charged by the Federal Reserve Bank to member banks is called the "rediscount rate."

Let us trace through a hypothetical case. Suppose X borrows \$1,000 for 60 days from the bank on his note, pledging as security part of the stock of goods of his business. The bank's discount rate is 6 per cent, so the interest for 60 days would be 1 per cent, or \$10 deducted in advance from the loan. X would then receive \$990, and in 60 days would repay \$1,000, leaving the bank \$10 profit on the loan. But suppose that owing to unexpected withdrawals the bank runs short of reserves. It could then

(if it liked immediately) rediscount X's note at its Federal Reserve Bank at the official rediscount rate. Suppose the rediscount rate is 3 per cent. Then X's bank will give his note to the Reserve Bank, receiving \$1,000 less 3 per cent annually ($\frac{1}{2}$ of 1 per cent for 60 days) rediscount, or a total of \$995 reserves by rediscounting the note (\$1,000 less \$5 interest charged by the Federal Reserve Bank). When X repays the \$1,000 at the end of 60 days, his bank will pay the \$1,000 to the Reserve Bank and the note will be returned to him. X has borrowed from his bank at 6 per cent; the bank has borrowed from the Federal Reserve Bank at 3 per cent. X's bank has made \$5 on the deal, and the Reserve Bank has made \$5, accounting for the total \$10 of interest X pays. X's bank would have made the full \$10 itself if it had not rediscounted the note; but when pressed for reserves it was able to rediscount the note and still make a profit on the loan since the rediscount rate was lower than its own discount rate to X.

Thus the Federal Reserve Banks may be able to control credit expansion by member banks through changing the rediscount rate. Suppose that in a boom period banks have been making large loans and investments, so that excess reserves are largely exhausted. At such a time the banks could keep on expanding loans and investments only by obtaining more reserves, which they might get by rediscounting loans. But by raising the rediscount rate to, let us say, 6 per cent, the Federal Reserve Bank could wipe out the profit on a loan. By doing this it would check credit expansion by its member banks, for they will continue to make new loans and investments only so long as it is profitable to do so. It is possible that the member bank will raise the interest rates charged to customers, in order to re-establish a margin of profit on rediscounted notes. But this higher interest rate to customers will itself check borrowing. In either case, the higher rediscount rate charged by the Federal Reserve Banks has acted to check bank credit expansion.

Take now the opposite case, where the Federal Reserve wishes to stimulate recovery from a depression by stimulating bank credit expansion. Suppose in such a case all Federal Reserve Banks lower rediscount rates to, let us say, 1 per cent. Will this stimulate bank credit expansion? Probably not, although it may have some effect. The reason is simple. In a depression, banks typically have large excess reserves lying idle because of bankers' inability to find what they consider safe and profitable loans and investments and because of the absence of persons and businesses wishing to borrow money for productive uses during depression. Therefore, if a bank makes a loan to a customer it would certainly not rediscount the note. By holding the note itself, the member bank earns the entire interest—it would be foolish to rediscount the note when it has already such large idle reserves. The lower Federal Reserve rediscount rate is of little significance to the member banker, for he does not wish to rediscount. The interest rate charged is of little significance to a person who does not want to borrow.

From this it is evident that higher rediscount rates may be fairly

effective in checking bank credit expansion if they come at a time when low excess reserves force member banks to rediscount. But low rediscount rates are much less likely to stimulate new bank loans and investments during depression periods, for when banks have excess reserves they do not have to rediscount. Lower rediscount rates inaugurated early in recession might aid in checking the contraction of bank loans and investments. In such a period, lower rediscount rates would make it easier for banks to obtain funds without liquidating assets. But a tradition against rediscounting except in emergencies has grown up over a long period among American bankers, and this tradition sharply reduces the effectiveness of lower rediscount rates in stimulating bank credit expansion.

"Advances"

A member bank may borrow from its Federal Reserve Bank on its own promissory note, just as a customer borrows from his bank on his promissory note. If it borrows on such a note, rather than rediscounting a customer's note, a member bank must pledge collateral as security for the "advance" made by the Federal Reserve Bank. This collateral may be anything acceptable to the Federal Reserve Bank (often government bills), but the rate charged on advances secured by collateral other than Government securities or "eligible paper" must be at least $\frac{1}{2}$ of 1 per cent higher than the rediscount rate. The comments just made about the effectiveness of changes in rediscount rates apply also to changes in the rate charged on advances. Higher rates when excess reserves are low have some power to check credit expansion; lower rates at any time are of very limited effectiveness because of the tradition against member bank borrowing, especially in times of large excess reserves when member banks have no need to borrow.

The Reserve authorities may also affect credit expansion through their policy on advances by altering what they consider "acceptable" collateral. If they wish to check member bank borrowing and hence member bank credit expansion, they may make stringent requirements as to the collateral presented by banks desiring advances. If they want to stimulate credit expansion, they may relax their standards of acceptable collateral. But Federal Reserve policy on advances has been of minor importance because of the very limited use that bankers have made of this device.

Direct pressure

When the Federal Reserve authorities wish to restrict member bank credit expansion, they may use "direct pressure," or "moral suasion." This pressure may take any one of several forms. Examiners may be instructed to become more stringent in their requirements for "good" loans and investments. Federal Reserve officials may send circular letters to member banks, pointing out the dangers of "speculative" loans and investments; and particular offending bankers may be singled out for personal letters or interviews. Federal Reserve officials may make

public statements warning against overexpansion of business and of loans to business for speculative purposes. In extreme cases, the Reserve Banks may simply refuse to loan to offending member banks—that is they may refuse to rediscount or make advances to the offending bank, thereby limiting its power to obtain reserves for further credit expansion.

Direct pressure may also be used in an attempt to persuade bankers to make more loans and investments in bad times. Federal Reserve officials may make public statements pointing out the need for business expansion and that excess reserves are available to finance such expansion; they may relax examiners' standards of "good" loans and investments; and so on. But such moral suasion has not generally proved very successful in controlling the volume of bank credit.

Changes in member bank reserve requirements

Before 1933 discount rate changes, open market operations and direct pressure were the important instruments available to the Federal Reserve authorities for regulating bank credit expansion. As part of the legislation of 1933, a more drastic power was given the Board of Governors—the power to change actual legal reserve requirements for member banks from the then existing levels of 7 per cent, 10 per cent, and 13 per cent for the three classes of banks. This power was enacted into permanent law by the Federal Reserve Act of 1935, which greatly centralized control of the Federal Reserve System in the hands of the Board of Governors. At present the Board has the power to vary legal reserve requirements between the old levels and levels twice as high—that is, between 7 per cent, 10 per cent, and 13 per cent, and 14 per cent, 20 per cent, and 26 per cent on demand deposits, and between 3 per cent and 6 per cent on time deposits. Up to 1942, the Board had exercised this power five times, raising reserve requirements in 1936 and twice in early 1937 as the danger of an inflationary boom on large excess reserves became serious, lowering requirements in 1938 in an effort to stimulate expansion during the business contraction following 1937, and raising them to the legal maximum in the war boom and inflationary expansion of late 1941. In 1942, as excess reserves of New York and Chicago banks were depleted by shifts of funds, the currency drain, and increasing required reserves for growing deposits, the Board lowered central reserve city requirements against demand deposits to 20 per cent, in order to permit these banks to continue to buy large amounts of government securities.

By raising reserve requirements, the Board is able directly to decrease member banks' excess reserves and thereby to restrict credit expansion. Suppose a member bank has \$1,000,000 demand deposits and \$200,000 reserves, the legal reserve ratio being 12 per cent. This leaves \$80,000 of excess over the required \$120,000 reserves. If then the Board raises the legal ratio to 14 per cent, this increases the reserve required against the same deposits to \$140,000, and decreases excess reserves to \$60,000. Similarly, if the legal reserve ratio were lowered, this would directly increase excess reserves.

How effective changes in legal reserve requirements will be in achieving the desired ends again depends on conditions prevailing at the time. Raising reserve requirements will be effective in checking credit expansion (or even in forcing contraction of credit) if it reduces excess reserves to levels the bankers consider undesirably, or "dangerously," low. In any case reducing excess reserves will have some limiting effect on credit expansion at the time, and the higher reserve requirements reduce the amount of loans and investments that can ultimately be made on existing reserves if banks should expand to this limit. Lowering reserve requirements will be effective in stimulating bank credit expansion only if bankers make new loans and investments on the basis of the excess reserves created by lowering reserve requirements. If bankers already had large excess reserves before the lowering of reserve requirements, the addition of still more excess reserves will not greatly stimulate lending and investing, though any increase in excess reserves puts pressure on bankers to make loans and investments in order to earn all possible interest on idle reserves.

Changes in "margin" requirements

All the credit control powers noted so far relate to the volume of bank credit in general, rather than to the particular use made of it. But the Federal Reserve Act specifically directs the Federal Reserve authorities to guard against the undue use of bank credit for speculative purposes. For the most part, reliance is placed on the judgment of the individual banker to see that the loans made are "sound" and not for speculative purposes, and bankers are continually subjected to supervision by bank examiners. But the Board of Governors has been given special powers to control the use of bank credit that may be used for speculation in securities.

Many times customers buy securities "on margin"—that is, by paying the broker a cash "margin" (down payment) and borrowing the rest of the purchase price from the broker, leaving the newly purchased securities with the broker as collateral for the loan. The brokers, in turn, typically borrow from commercial banks the funds that the security buyer does not put up as margin (cash). The smaller the margin required, the more speculators can borrow of the purchase price of the securities bought; the higher the margin, the less they can borrow.

Therefore, if margin requirements are raised, the use of bank credit for speculation in securities is to some extent checked. If margin requirements are lowered, it becomes easier to use credit in buying securities. In the Securities Exchange Act of 1934 the Board of Governors was given power to set minimum margin requirements for dealings on the major securities exchanges. This power to change margin requirements has been exercised by the Board several times since 1934, with margin requirements ranging from 40 per cent to 55 per cent at different times. These requirements are substantially higher than those in effect during the stock market boom of the late 1920's.

Although speculative activity in securities has played an important part in most booms, this power to change margin requirements is of limited value in maintaining business stability. This is true for two reasons. First, control of the use of credit in security markets, even if it could be made completely effective, would touch only a small segment of the economy, and the general control of the total quantity of credit is of much broader importance. Second, higher margin requirements are only a partial check against "undue" use of credit in security speculation. Higher margin requirements can check direct extension of bank credit through brokers to customers for use in margin purchase speculation, but they cannot prevent the use of other bank credit for this purpose. If Mr. A borrows \$10,000 directly at his bank (on his personal note), he may use this to pay the broker in full so that higher margin requirements are of no effect. Yet he is still using credit to speculate in securities. The power to change margin requirements is of considerable importance in the particular area of the security markets, but of much less importance than the overall credit controls for the economy as a whole.

Control of installment credit

During the 'twenties, one of the factors contributing to the steadily increasing consumer demand for goods was the rapid growth of installment credit. An increased extension of such credit by sellers has the same effect on total consumer buying power that a corresponding increase in bank credit extended to consumers would have. It might be thought that installment credit merely constitutes a postponement of payment, but so long as sellers continue to extend new installment credit as fast as the old is paid off, the increased total volume of credit remains outstanding; and during the 'twenties sellers not only replaced maturing installment credit but continually increased the total volume outstanding. With the depression, such credit was curtailed in the general process of liquidation, and as purchases were paid off without new extensions of credit the total volume of installment credit was contracted, thereby reinforcing the general process of credit and business contraction and liquidation.

With the upswing of business activity following 1933, installment credit again became a major factor in consumer buying. As long as the major problem was to increase spending and employment, this increase in the volume of installment credit was consonant with general credit aims, but when, in 1941, the limits of productive capacity, employment, and raw materials were reached in many industries crucial to the defense effort, it became necessary to divert the available production from civilian to defense uses. Although it was not yet considered necessary to impose drastic overall credit controls, it was thought advisable to curtail consumer spending on such articles as autos, refrigerators, radios, furnaces, and so on, which use large quantities of essential defense resources, in order to prevent sharp price rises that otherwise would have resulted from sharply restricted production (imposed by priorities) in the face of unrestricted consumer demand.

The Federal Reserve was chosen to impose and administer this restriction of installment credit in August 1941. The original regulation applied only to certain articles directly competing with defense production and to cash loans on an installment basis under \$1,000, prescribing minimum down payments and maximum length of payment period. The original regulation was very mild, but the breadth and stringency of the control have been increased as the urgency of the problem has increased.

Installment credit control is a form of selective credit control somewhat similar to the Board's control over margin requirements in the securities markets. Like the control over margin requirements, it fulfills an important function in a special area, but is of comparatively small importance as an overall credit control device. As with margin requirements, regulation of installment credit practices is closely involved with general social pressure for "reform" of "excessive" practices. Whether such controls will continue to be exercised after the war emergency that called them forth has passed is a matter that only time can tell.

Credit Policy in Wartime

In 1942 a serious dilemma arose for the Federal Reserve authorities. In spite of the staggering increases in war expenditures, taxes were increased only a fraction of this amount and purchases of government securities by individuals and nonbanking institutions fell far short of providing the additional funds needed. Thus it soon became apparent that probably over half of the entire new public debt issues would have to be sold to the banks. For fiscal 1943, it was generally estimated that the banking system would take over \$20,000,000,000 of the total increase of about \$60,000,000,000 in the public debt, as compared to only slightly over \$20,000,000,000 of federal tax collections.

Coupled with decreasing supplies of civilian goods for sale, this tremendous increase in expenditures through credit expansion exerted a powerful inflationary pressure, greatly increasing the difficulty of maintaining the direct price controls established by the Office of Price Administration. Owing to increasing reserve requirements against soaring deposits and to a heavy drain of currency into circulation, member bank excess reserves were rapidly depleted. Federal Reserve officials were then faced with the problem—whether to permit exhaustion of excess reserves to check the inflationary credit expansion, or to provide added excess reserves to enable the banks to continue to buy vast amounts of securities, in spite of the inflationary consequences. They chose the latter alternative, a course made virtually inevitable by the failure of Congress to enact an adequate program of taxation and forced saving. Early in 1942, the Reserve Banks began to buy heavily in the open market to provide additional reserves, and early in the autumn the Board lowered reserve requirements on demand deposits at central reserve city banks by three steps to 20 per cent, equal to the reserve city requirement. Bank deposits and holdings of government securities rose at an unprecedented rate, deposits in 1943 approaching a level double that of the 1929 peak.

Most drastic inflations have come in wartime or shortly thereafter. Yet under war conditions, credit policy is largely dictated by war needs. Government expenditures and taxes, outside the control of Federal Reserve authorities, dominate the picture, with taxes grossly inadequate in virtually every case in recent history. At no other time are the limitations on the power of credit control alone to check inflation more clearly exhibited; credit policy in its broader aspects is essentially determined by the volume of war expenditures relative to the volume of taxation plus borrowing by the nonbanking public, since the central bank must assure the taking of government securities not sold to the public. Thus, as will later be considered more fully, under the dangerous inflationary circumstances associated with war conditions, credit policy must be secondary in importance to fiscal policy and to other governmental direct actions in promoting economic stability at desired levels.

Summary

By reducing excess reserves and by making it more difficult for member banks to obtain additional reserves, the Federal Reserve authorities may rather effectively check credit expansion if they are willing to take sufficiently drastic steps. But the Federal Reserve's power to check credit expansion is limited by its power to exhaust member banks' excess reserves. With the recent excess reserves of \$6,000,000,000, for example, the Federal Reserve would have exhausted all its securities in open market sales long before excess reserves were seriously impaired. Likewise, even though, in addition to the open market sales, member bank reserve requirements were raised to the Board's limit of 14 per cent, 20 per cent, and 26 per cent, this would have left excess reserves of some \$2,500,000,000 to serve as the basis for credit expansion. Since member banks would not need to borrow from the Federal Reserve Banks under such circumstances, higher rates on advances and rediscounts would be ineffective. Direct pressure would be available, but this has never proved a major instrument of control. Higher margin requirements are useful only in the fairly narrow area of the securities markets, and installment credit control was not then available. The Federal Reserve authorities could only hope for additional power to raise reserve requirements further in case of an inflationary price rise (and they did in fact request this additional power from Congress); the situation was indicative of the limitations on Federal Reserve powers. And in fact it was the rapid drain of currency into hand-to-hand circulation in 1940, 1941 and 1942, a phenomenon largely outside the control of Reserve authorities, that was far more effective in reducing excess reserves than anything the Federal Reserve could have done to achieve this end.

By increasing excess reserves and making it easier for member banks to obtain more reserves, Reserve authorities may attempt to stimulate bank credit expansion. But there is no way of forcing the banks to make loans if bankers do not feel that safe and profitable loans and investments are available or if businessmen and farmers do not want to borrow. The

authorities can make reserves available to member banks in the hope that this will lead to expansion of loans and investments, but they cannot insure a demand for bank loans or a willingness on the part of bankers to make loans.

Establishment of the Federal Reserve system and especially the centralization of authority provided by the Act of 1935 have greatly increased the possibilities of control of bank credit expansion, but these possibilities of control are still very limited, especially in regard to checking contraction of loans and investments in times of business recession and stimulating expansion in depression. The quantity of credit (bank checks), which makes up by far the largest part of our circulating medium, is still largely dependent on the loans and investments extended or called in by the commercial banks—Federal Reserve control has only partially removed the dangers of drastic fluctuations in the amount of bank credit extended. Lastly, even though control of bank credit expansion and contraction could do much toward mitigating business fluctuations through controlling bank credit extension, one must not overestimate the possibilities of controlling business fluctuations through controlling the extension of bank credit. Booms and depressions have many causes and many complex characteristics, only some of which are susceptible to control by monetary means.

Part IX

GENERAL BUSINESS FLUCTUATIONS AND ECONOMIC POLICY

PREAMBLE

NO ECONOMIC question of the last decade has occasioned so much bitter dispute as that of how to “cure” the long depression of the 1930’s—a very proper question in view of the severe economic disruption and social distress wrought by this worst of all modern economic debacles. Unfortunately, booms and depressions are exceedingly complex economic phenomena, and not even professional economists, much less the lay public, have been able to agree on the best way to go about straightening matters out. “Spend more!” cries one group. “Spend less!” shouts another. “The thing to do is get prices and costs back into line,” insists a third.

Upon so complex a question one must be content with a good deal less than “final answers” in an elementary course—indeed, in any course, if the conflicting opinions of the experts are any indication. Yet much can be said about the effects of different policies on business fluctuations, and the way to begin is by explaining some of the major factors at work in such oscillations.

CHAPTER 44

General Business Fluctuations

What Are "Business Cycles"?

THE most elementary survey of the economic history of industrialized countries shows clearly recurrent booms and depressions. Therefore it has become common to speak of "business cycles" or of "the business cycle." Strictly speaking, there is no such thing as "the" business cycle. Although there have been rather regularly recurring waves of business expansion and contraction for the last century and a half in the industrialized countries of the world, these waves or fluctuations have been only roughly similar. To speak of "the" business cycle is to give a definitely misleading connotation. It may even be preferable to avoid the word "cycle" since cycle connotes a regularity that business fluctuations have not always had. There have usually been minor business fluctuations (cycles) about every three or four years, and larger fluctuations on an average of about every nine years. Yet these periods have varied widely, both as to amplitude and frequency. For example, the "depression" following 1929 was more severe than any other in American history, and longer than any other except possibly that following the crisis of 1873. But during the period since 1929 there have been two minor "cycles," superimposed upon the major cycle. There was a revival in late 1933 and early 1934, followed by a recession; and then a very sharp upswing in 1936 and early 1937, followed by a severe recession. By early 1941 business activity had again achieved high levels with industrial production on the highest level ever, well above 1929. Yet there was still an "army of unemployed." This brief period illustrates well these two major characteristics of business cycles—first, their fairly regular recurrence, and second, the fact that every "cycle" is different from every other.

In spite of such differences, it is easy to point out certain major characteristics that are present in every cycle. Indeed, it is useful in studying them to set up a rough "pattern" of business fluctuations, including in the pattern those characteristics that are most common and important for all cycles. But it must be remembered that this is only a "pattern," and that an explanation of any particular fluctuation must take account of the many characteristics peculiar to it and may require modification of the basic pattern.

Although each business cycle in history has had its own individual

characteristics, the essence of all of them has been the fluctuation in (1) *employment* and (2) *production*. The essence of a "depression" is a low level of physical production of goods and widespread unemployment of productive resources, especially labor. The main characteristics of what we generally call "prosperity" are relatively full employment and a high level of productive activity. A third major characteristic, following from these two, is fluctuations in *income*. Fluctuations of "real income" are the same as fluctuations in production, but there also occur fluctuations in total money income being paid out for the use of productive agents, fluctuations that play an important role in the cyclical process. A fourth major characteristic—and one which does not necessarily follow

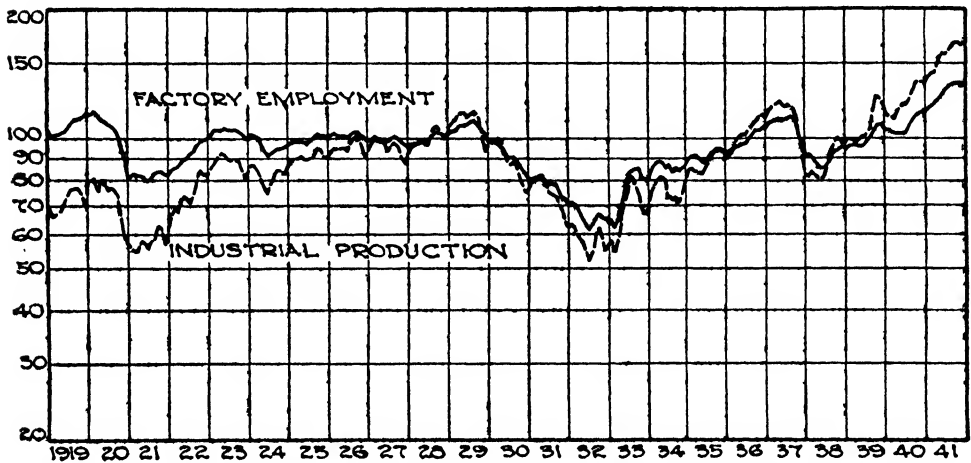


Fig. 44—1. Production and employment, 1919–1941 (1923–25 = 100).*

* Federal Reserve Board Index of Industrial Production; U. S. Bureau of Labor Statistics Index of Factory Employment.

from the first two above—is fluctuations in the *price level*. All American business cycles except that culminating in 1929 have been marked by a rising price level in revival and prosperity and a falling price level in recession and depression. And although the the prosperity of the late 1920's occurred without a rising price level, there never has been a serious depression without sharply falling prices. The size of the fluctuations in these elements has varied considerably in different cycles, but they are the central elements. Figure 44—1, showing cyclical fluctuations in industrial production and factory employment over the last two decades, gives a general idea of this cyclical pattern. Inclusion of very recent years would show a tremendous rise under the stimulus of the war.

Although people generally think of "the business cycle" or of "business cycles," there exist also individual cycles of prices and production for individual phases of activity—for example, hog cycles, building cycles, ship-building cycles, and so on—which do not always coincide with the general business fluctuations. Often it is very difficult to tell when the

general cycle turns come, because many series turn up or down before others. The point selected as the turn of the cycle is arbitrary within fairly wide limits. (For example, whether the upturn after the depression of 1929-32 came in 1932 or after the bank holiday and financial measures of March, 1933, is still debated.) Some writers argue that the general business cycle is nothing but a coincidence of many of these individual cycles. This coincidence exists, of course, but the problem is to explain why the fluctuations in different lines of business are synchronized so as to cause such severe fluctuations in employment and production.

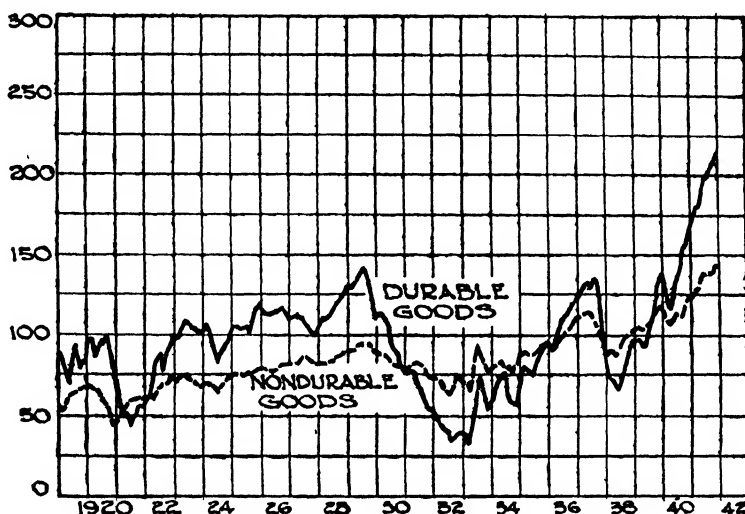


Fig. 44—2. Production of durable and nondurable goods
(1935-1939 = 100).*

* Reproduced from "The New Index of Industrial Production," Board of Governors of the Federal Reserve System, p. 10. Indexes adjusted for seasonal variations.

Fluctuations in production and employment come largely in the capital goods and durable consumers' goods industries. Cyclical fluctuations are much sharper in such "capital goods" industries as steel, electrical machinery, and locomotives, and in such "durable consumers' goods" industries as building construction and automobiles than in shoes, clothing, foods, and other such relatively nondurable consumers' goods. It has been estimated that capital goods industries giving only about 1/10 of the total employment in 1929 were responsible for almost 1/3 of the total unemployment in 1933. This does not mean nondurable consumers' goods industries are not affected by cyclical fluctuations, but rather that the heart of the problem lies in the capital and durable consumers' goods industries. Booms center there, and revival must come largely through increased investment in these areas.

The primary reason for this difference between the behavior of durable and nondurable consumers' goods is fairly obvious. As between food-stuffs and overcoats, for example, in a depression it is fairly easy to wear

one's old overcoat another year but if one has any income at all he is likely to keep up fairly well his consumption of foodstuffs. If it is a matter of buying a new washing machine, the purchase is still more likely to be postponed. For a new house, the deferment is easier yet. Purchases of "necessities" are maintained if possible while purchases of "non-necessities" tend to be postponed. The more durable a good is, the less likely is its replacement to be "necessary" in any given year, though eventually the replacement of durable goods becomes more and more "necessary." And exactly the same principle holds for decisions of entrepreneurs to invest in new machinery, buildings, and other durable equipment. Figure 44—2 shows how much more violent have been the fluctuations in production of durable than in production of nondurable goods over the past two decades. The significance of the distinction made above between capital goods and consumers' goods will appear later.

Entrepreneurs and the Level of Business Activity

Under a private enterprise, capitalistic type of economic organization, such as has existed in the United States, the decisions as to the level of production and employment are made *primarily* by entrepreneurs, whose major aim is the maximization of money profits. These entrepreneurs, of course, may be either individual proprietors or those responsible for management of corporations—at present corporate entrepreneurs control the largest part of the production and employment in the country, excluding agriculture. Governments (federal, state, and local) also directly control the level of employment to some extent—partly through employment of such regular workers as clerks, mailmen, and janitors, and partly through such extraordinary employment-maintaining activities as WPA, PWA, and so on. However, the former type constitutes a small portion of total employment; and the latter type comes largely as a result of business cycles, since if there were no depression there would be little need for such government "work-relief" programs as the WPA. Therefore, government activities are disregarded here in the preliminary examination of the causes of business fluctuations.

If an entrepreneur can increase his profits by doing so, he will increase output, which generally involves hiring additional laborers, buying more raw materials, building or buying new plant and machines, borrowing funds. If he finds that at his existing output he is incurring losses that he can reduce by laying off agents and decreasing output, he will do so. Thus, if a seller's revenue opportunities remain the same and costs of productive services rise, the marginal value returns on agents employed will no longer be sufficient to cover the marginal costs of employing so many agents, and agents will be laid off until the maximum profit or minimum loss position is again reached. Gradually firms will withdraw from production if losses continue. Conversely, if demand conditions improve while costs remain stable, it will be profitable to hire more productive agents and increase production: and a similar result obtains if demand conditions remain unchanged while costs fall (for example,

because of technological improvements, as was the case during the 1920's). Each entrepreneur strives to maximize his profits, and the fluctuations in output and employment that constitute business cycles come largely as the result of many such entrepreneurial decisions to hire more or less agents, to increase or decrease output, in response to changing anticipations of revenue-cost relations—decisions that are more or less synchronized by the cyclical forces at work.

In such a complicated economy as ours, where production is roundabout and for a future market, it should be obvious that decisions of entrepreneurs must depend largely on their expectations as to future conditions. Especially in cases where production requires the use of expensive, heavy, durable machinery and equipment are expectations for the future important. Here the entrepreneur's profit calculations must run over the whole life of the fixed investment (for example, the life of a machine). And it is in precisely such heavy, capital goods industries that cyclical fluctuations are greatest. *It is the expectations of entrepreneurs as to future revenue-cost relations that primarily determine the level of investment in such industries, and hence primarily determine the stage of the "business cycle."*

In view of the long period over which profit calculations must run on many investments, entrepreneurs are inevitably confronted with a considerable amount of uncertainty. Whether each entrepreneur's profit expectations are favorable or unfavorable depends largely on how he evaluates and reacts to this uncertainty. Most entrepreneurs, if they are very uncertain as to future conditions, are probably loath to make heavy investments. For many people, such uncertainty may be as important in holding back new investment as is outright pessimism, and in fact it is often difficult to separate the two. Expectations depend on the vast complex of continually changing facts, hopes, and suspicions that make up the economic, social, and political environment at any time. And to explain the level of business activity requires an explanation of why the expectations of entrepreneurs are what they are—why entrepreneurs decide to increase or decrease output and employment; in large part why they do or do not undertake large scale new investments.

The Framework of the Cycle

Some writers have attempted to build business cycle theories on some one basic cause. Any such simple explanation of such a complex phenomenon should be viewed with much scepticism. Today, although many outstanding writers emphasize certain factors as most important causally, all agree that any acceptable explanation must take into consideration a considerable number of factors. These factors may be separated into two major types—first, certain characteristics inherent in business fluctuations that tend to lead on to the next phase of the cycle; and second, "extraneous" factors, such as wars, crop disasters, political activities, and so on, which affect the level of business activity "from without." History gives support to the emphasis on both types of factors. Since business fluctu-

ations have been surprisingly regular for well over a century, not only here but in other industrialized countries, it is reasonable to infer that there are certain characteristics of the fluctuations that tend to perpetuate themselves. On the other hand, there is no certainty that they would be self-perpetuating, since there have always been extraneous forces at work. For example, there is no doubt that gold discoveries and the Spanish-American War were very important in bringing revival and prosperity in the late 1890's. Whether the revival would have come as soon, or at all, if it had not been for these events, no one knows. In the brief discussion in this chapter, extraneous forces are given little specific attention, since each is a separate case, but this neglect in no way implies that they are not important. They may reinforce the cyclical movement, partially counteract it, or even completely reverse it. But out of the welter of conflicting emphases on different inherent and extraneous factors there is one major point of agreement. This is that the part played by the monetary system (including the banking system) is of basic importance in explaining the cyclical process, although some writers speak of monetary changes as "causes" and some as "necessary reinforcing factors" for the development of serious booms and depressions.

In considering cyclical business fluctuations, it is useful to trace through very briefly one "cycle" in preliminary fashion, pointing out certain important and characteristic factors in each phase. In each phase certain major factors that tend to lead to the following phase—from depression to revival to prosperity to recession (possibly preceded by a temporary "crisis") and back to depression—are indicated. There is little doubt that once revival is well started it tends to perpetuate itself—as does also recession. Moreover, it is fairly easy to see why booms do not continue forever. But as to whether there are any forces in depression that necessarily lead to revival is less certain. Division of the cycle into four phases is merely for convenience in discussion. There is no real dividing line between phases; cyclical fluctuations are a continuously developing and dynamic phenomenon.

We may begin with depression. The major characteristics are a large amount of unemployment and a low level of production (primarily because of a low level of investment in capital and durable consumers' goods industries), which implies of course a low level of income. Now suppose for some reason (let us say, a war abroad or the introduction of an important new product) entrepreneurs decide there is profit to be made by expanding output, using existing plants, or even building new ones. Expanding output will mean taking on more workers, buying more raw materials, possibly building or buying new machinery, all of which mean increased spendable incomes for sellers of productive services (especially laborers). Other entrepreneurs, seeing this increased productive activity and finding increased sales resulting from increased incomes, are likely to increase output and make plans for improvements in plant. If other entrepreneurs undertake new projects (increased output, new investments, and so on), this in turn further increases incomes and money demand for

goods. Even small increases in sales will deplete existing small inventories and make replacements necessary, and it is likely that replacements will be larger than the sales made in view of increasingly favorable expectations. People who have been holding large idle cash balances are likely to reduce the size of these idle balances since the future no longer looks so uncertain—and banks become more liberal in making loans and investments. This monetary expansion must have one of two effects—either prices will rise or more goods will be sold at the same prices, probably a combination of both. Expanding consumer demand provides a continued stimulus to increased production and new investments. Central in the cumulative expansion are the bettered profit expectations of entrepreneurs, leading them to expand output and undertake new investment opportunities. As the revival continues, revenues rise faster than many costs, giving wider profit margins and more incentives to increased output and investment. Costs typically lag because (1) there are many unemployed agents so that entrepreneurs can get more workers, funds, and so on, without much increasing agent payments, and (2) many costs are fixed for considerable periods in money terms (for example, salaries and interest on bonds). Because of this lag, such costs are often termed “sticky” costs.

Gradually, however, operating costs will rise faster as production expands; and such semifixed costs as rents, salaries, and so on, also begin to move. The depression “backlog” of unsatisfied demands is gradually filled as new investments and production are undertaken, and other factors help cut down profit opportunities. As costs begin to rise faster than revenues, profit margins are decreased and entrepreneurs become less inclined to make expansive plans for the future. If cost-revenue relations become sufficiently unfavorable, entrepreneurs will not only cease expansion but will decrease output and lay off workers. Thus the essence of the upswing is two-fold—first the expansion of monetary demand (for consumers’ and producers’ goods) with the revival of investment and production activity; and second, the existence of sticky costs so that monetary expansion and rising prices widen profit margins and stimulate expansion. The end of the expansion comes when entrepreneurial expectations, for any of a variety of reasons, become unfavorable, leading to stable or decreased output and employment.

The downswing of the cycle is essentially the same process in reverse—decreased monetary demand, coupled with sticky costs, is at the center of the contraction process. When monetary demand falls but costs remain high, decreased or negative profits result, leading entrepreneurs to curtail output and drop plans for new investment. These acts in turn further decrease income and monetary demand, and this contraction is intensified by loan and investment liquidation by the banks, which increases losses and makes prospective investment still less attractive. Here again it is lagging costs that intensify the effects of contraction of monetary demand. This pattern of cyclical fluctuations of monetary demand against sticky costs as primary determinant of shifting entrepreneurial expectations and

behavior is greatly oversimplified, but it should be borne in mind throughout the following pages as the core of cyclical fluctuations.

Depression

A more detailed examination of the factors that may lead to revival from depression may now be undertaken.

In depression unemployment is large and most wage rates have fallen to the point where they are again roughly in line with depressed selling prices. Other costs have likewise fallen. Prices of raw materials are low, and the cost of borrowed funds (that is, interest rates) is likewise very low, although loans are available only to those who have excellent security and general credit standing. Having large excess reserves, the banks are anxious to loan to "good" risks. Even such semifixed costs as long-term interest charges, depreciation and rents have been reduced somewhat, especially if the depression is a severe one in which creditors are glad to compromise to assure themselves of some income rather than suffer complete debtor default. Thus costs have been more or less brought into line with the lower selling prices, permitting businessmen to make "new soundings" as a basis for expanding operations.

During the recession and early depression the banks brought strong pressure for repayment of all loans not strictly first class, thereby helping to force the elimination of weak firms from the business scene. The remaining firms are borrowing few or no funds from the banks, so that bankers find themselves with large excess reserves and are anxious to make loans and investments, but only to strictly first-class risks. What constitutes a first-class risk depends, of course, largely on the expectations of the banker as to future business conditions; if revival once gets started and favorable expectations begin to replace pessimism and uncertainty a large supply of loanable funds becomes available at the banks. Liquidation of private and business indebtedness during the period just passed has placed most remaining firms, as well as the banks, in strong cash positions.

As a result of the process of liquidation, inventories of almost all businesses are very low. This means that any increase in sales will quickly deplete existing stocks, giving rise to a replacement demand of considerable proportions, for an increase in sales is apt to mean improved expectations so that larger inventories will be desired for the future.

In depression there is a low level of total money income—less money is being paid out for people to spend and invest. Decreased consumption has been largely in the lines of postponable expenditures, such as automobiles and new houses, as people simply make the old car, house, overcoat, or whatnot, last a little longer. But as such postponement continues, there tends to be accumulated a large potential demand for durable goods. If purchasing power (consumers' income) revives, people will be anxious to replace their worn cars and overcoats with new ones. Thus, while in the depression the demand for durable consumers' goods falls off more sharply than that for nondurables, the former typically has

a much sharper increase when revival comes. This, roughly, is what accounts for the fact that there are always much greater fluctuations in output, employment, and prices in the durable than in the nondurable goods industries.

This principle applies in capital goods industries with special intensity. Entrepreneurs, like consumers, tend in depression to defer postponable investments (for example, decide to use old machines another year before replacing them, even though replacement this year had been planned). When revival comes, entrepreneurs find themselves faced with the necessity of replacing a great deal of machinery, buildings, and so on, that should ordinarily have been replaced a year or two or three sooner.

Moreover, there is another reason why the demand for productive agents (for example, machines to make shoes) fluctuates more sharply than does the demand for the ultimate product (shoes), a tendency which has been called the "acceleration" or "derived demand" principle. Suppose, for example, that there are 10,000 shoe machines used in the shoe industry, producing 1,000,000 shoes per year, so that each machine produces 100 pairs of shoes annually, and that the average life of each machine is 10 years. Each year 1,000 machines wear out and have to be replaced, so that the demand for new shoe machines is 1,000 per annum. Suppose further that as revival gets under way the demand for shoes increases to 1,100,000 at existing prices, an increase of 10 per cent. This means that 1,000 new machines will be needed to make the additional 100,000 shoes, in addition to the regular 1,000 machine replacement demand. This year the demand for new shoe machines will be 2,000, an increase of 100 per cent owing to an increase of only 10 per cent in the demand for shoes.

The principle works in the same fashion at the downturn of the cycle into recession. Suppose there is a 20 per cent decrease in demand for shoes from 1,000,000 to 800,000 pairs at existing prices. Then only 8,000 machines will be needed to produce the shoes—not only will there be no new machines bought (a 100 per cent decrease) but there will be 1,000 idle machines left in the industry throughout the year. Unless the situation changes, again the following year there will be no new machines bought. Thus comparatively small fluctuations in the demand for the final product may cause very large fluctuations in the demand for the factors producing the product, though the illustration used is a purely hypothetical one. In general, the more durable the good the greater is the force of the derived demand relation, for the more durable the good the smaller will be the proportion of the total supply replaced each year. This same principle can be applied to such durable consumers' goods as houses and automobiles. For example, if houses ordinarily last fifty years so that each year $1/50$ of all houses are replaced (say, 10,000 annually) and owing to revival 10,000 young couples who had been living with relatives decide to build new homes, this increase of $1/50$, or 2 per cent, in the demand for homes will result in a 100 per cent increase in the demand for new homes. Thus a comparatively slight increase in demand at the beginning of revival

may lead to important effects on production and employment in industries that reflect the acceleration principle, a group including a majority of the heavy industries, which suffer most in the depression.

Since cyclical fluctuations in production and employment are primarily centered in capital goods and durable consumers' goods industries (partly because of the principle of derived demand and partly because of the postponability of durable goods replacements), cyclical fluctuations show up primarily in investments undertaken in such industries, including investment in inventories of goods on hand but not yet sold. Figure



Fig. 44—3. Net capital formation, 1919–1935 (in billions of dollars). *

* Data from Simon Kuznets, *National Income and Capital Formation, 1919–1935* (National Bureau of Economic Research, 1937), Table 13, p. 48. Figures are expressed in 1929 prices. (Used by permission of the copyright owner.)

44—3 shows the drastic fluctuations that have taken place during recent years in the net investments made, including investment in unsold inventories. During the worst years of the depression of the 1930's, net capital formation was actually a negative figure—that is, businessmen, farmers, and consumers did not even replace wear and tear on plant and equipment, much less make new investments. Part of this disinvestment came from selling off inventories and failing to replace them. However, never before had there been such net disinvestment in a depression.

One last basic characteristic of depression, implied throughout, should be separately mentioned. This is the general attitude of pessimism and uncertainty which prevails, reflecting the entire social, economic and political complex of the times. This *weltanschauung* hinders any prospects of revival in at least three ways. First, it helps keep consumers

from increasing their rate of expenditures and prompts them to postpone buying as long as possible. Second, it makes entrepreneurs extremely reluctant to undertake new projects or to expand operations on old. In many cases expansion would involve borrowing funds, and the entrepreneur hesitates to undertake this financial responsibility in the face of such uncertainty. If he expects demand conditions to be even worse in the future, he naturally hesitates to undertake production of goods that he fears could be sold only at a loss, even though present cost-price relations may be profitable. Third, it makes bankers and savers extremely reluctant to part with their funds. Although interest rates are very low, loans are made only to very high-grade credit risks, and such credit risks seldom want to borrow in depression. Pessimism and uncertainty as to the future thus keep both entrepreneurs and investors from moving forward. In spite of generally favorable conditions, as long as this attitude prevails there is no one to start the ball rolling and to get revival under way.

Revival

Revival may be instigated by some fortuitous event, such as exceptionally good crops, important new inventions, or good international news; or it may come as the more daring entrepreneurs see and exploit the profit opportunities in particularly favorable lines, thus far left unexploited. Suppose that the stimulus is a group of new inventions, let us say comparable to the coming of the radio, the airplane, or the automobile, as has often been the case. As these industries expand plant and increase output, this stimulus may be expected to spread to other sectors of the economy in two interrelated ways. First, the funds paid out by the new industry for wages, raw materials, and as other costs will be spent in turn by the recipients, thus increasing the demand for numerous other products and stimulating other industries. Second, other businessmen and lenders, seeing the success of the new enterprises, will begin to wonder whether they have not been unduly hesitant and will look more favorably on possibilities of action on their own part. With the stage set for business revival, unless some particular offsetting event takes place, this original stimulus may well prove cumulative. As workers go back to work, they not only receive larger incomes to spend but they are more likely to spend the funds they receive. As new projects are undertaken, the demands for raw materials, for machinery, for parts, even though they are fairly small, represent large increases over the depression levels of inactivity, and this improvement in turn puts more funds into circulation and helps dissipate uncertainty and unfavorable expectations. As prospects for the future become visibly brighter, more and more entrepreneurs and investors feel that it is wise to "get in on the ground floor," and bankers begin to relax their stringent requirements on making loans and investments. The banks, having large excess reserves, are anxious to loan, and with brighter prospects for the future, loans that previously were unacceptable automatically move up into the

acceptable class. As consumer demand increases, the principle of derived demand acts to intensify the demand for producers' and consumers' durable goods.

In speeding the revival phase, the monetary factor is of central importance. Two primary sources of monetary change are at work—first, an increased velocity of circulation; and second, the creation of new credit by the banking system. Both expansionary forces increase cumulatively once revival gets under way, and it is this expansionary monetary impulse that accounts for a large proportion of the betterment in business activity, although the monetary changes may be to a large extent the result of other factors affecting consumers', entrepreneurs', and investors' actions.

With increases both in the quantity of money and credit and in velocity of circulation, more money is offered for goods and services, and this fact, coupled with lagging costs, means increased profit margins and increased production and employment. Profits often turn out to be more than expected, especially where output expansion means unit cost reductions through fuller utilization of existing fixed plant, a factor likely to be especially important in heavy goods industries. Good times and success further increase the rate of spending and better the expectations of consumers, business entrepreneurs, and bankers, the last group still anxious to make sound loans on their excess reserves.

Historically, many booms have centered around the development of one or a few important new industries. For example, during the middle of the nineteenth century railroad building, with the resultant demand for iron and steel products, was especially important. During the 1920's, the great expansion of the auto industry and the accompanying highway building program were very important. Today many economists feel that the reason for the great amount of unemployment from 1932 to 1941 centered in the absence of any such important undeveloped industries in which heavy investment might profitably be made. Some writers go further and argue that the United States has become a "mature" economy, in which profitable investment opportunities are increasingly scarce. To the extent that this argument is true, recovery from depression is much less likely to come than has previously been the case except through some external stimulus such as war or government spending. However, this position rests on the assumption that the recent relative absence of large-scale profitable investment opportunities is permanent, and this argument, although it carries some force, appears unduly pessimistic.¹ It is a common error to think of investment opportunities as something absolute rather than in their actual dependence on the relation between anticipated investment costs and revenues.

¹ In 1870 no less an authority than John Stuart Mill, the most influential economist of the period, wrote that the possibilities of capitalistic enterprise were substantially exhausted. Similar statements have been prevalent in the other severe depressions of history, especially during the long depressions of the 1870's and 1890's in this country, and again during the past decade.

Once a few entrepreneurs have braved the uncertain paths of introducing new products and new methods and have successfully undertaken new investments, there are many others ready and eager to follow. Once new projects have been successfully consummated, either in old or new products, investment in other fields is likely to follow, both because of the increased incomes put into circulation by expenditures on the new investment and because of the psychological effect on hesitant investors.

The process of revival, once started, tends to spiral upward in a cumulative process. Human beings are notoriously optimistic, once they get well started, and at no time is this characteristic more easily observable than in the revival and boom phases of the cycle. Roseate expectations for the future prompt new investments, both in expansion of present industries and in the development of new industries and processes. In the securities (stock) markets this optimistic tinge becomes especially evident, with prices of securities being bid up sharply on the basis of improved profit expectations for the issuing companies. Prices, rising under the influence of increased spending, tend to pull further away from lagging costs, giving additional stimulation to expanded production and employment. In turn bankers are more liberal with loans, investors more liberal with their funds, individuals more inclined to spend their incomes promptly, and so on. The demands of businesses to build up inventories that had been drawn down very low during depression constitute an important expansionary factor, especially since with restored faith in the future it becomes wise to buy now to "beat the price rise." This incentive to buying applies to all, individuals, merchants, manufacturers, and tends further to speed the upward movement of prices, for such concerted buying causes the very price rise that it is intended to "beat."

The central factor of this entire process of revival, leading into full-fledged prosperity, is the cumulative increase in money demand for goods, arising both from the decrease in the propensity to hoard and from the increase in bank loans and investments and other forms of credit. This increase in monetary demand is inseparably tied up with nonmonetary factors, partly as cause and partly as result, *but without the increase in monetary demand the revival could at best be only a very weak and abortive affair, completely incapable of rising cumulatively into full-fledged prosperity.* The part played by the banking system, as well as that played by changes in velocity, is therefore thrown into a prominent position. With such great powers of credit creation, the commercial banks of the country may exert tremendous pressure to stimulate and reinforce business revival, once it is started and once there is a substantial demand for loans. Without a banking system of this sort, it is inconceivable that revival and boom periods could have been as sharp and speculative as they have been or, what is perhaps more important, that recession and depression periods could have been so severe and painful. For the credit that the banks have created they can (and do) destroy very quickly and painfully when the peak is past and the process of deflation sets in.

Prosperity

Prosperity is an indefinable state of affairs, generally recognized by everyone but almost impossible to define specifically. For example, by late 1940 we had attained a high level of business activity, sharply up from the preceding lows, with practically all the characteristics of general prosperity except for the major exception that there remained some 8,000,000 unemployed. Was this prosperity? Or was it still depression? Or was it revival? There is no unarbitrary way of separating revival from prosperity. The 1940 period is a good example of the fact that each boom and depression must be considered to a substantial extent as an individual case.

As revival culminates in prosperity, three basic characteristics become apparent: first, productive activity is on a very high level, especially in the durable consumers' and producers' goods industries; second, employment has reached relatively high levels, great increases having come in the durable consumers' and producers' goods industries, which account for so much of the unemployment during depression; and third, total monetary income has reached a level far above the depression lows. But the picture is not as simple and roseate as this.

As revival progresses, the factors conducive to continued increase become less and less strong and the "vulnerability" to possible factors leading to a downturn greater and greater. The incentive to increased business activity during revival comes in considerable part because of the spread between costs and rising prices. "Profitable" investment opportunities develop everywhere in such a period. Investors and entrepreneurs naturally take advantage of the best opportunities first, so that as revival progresses recourse has to be had to less promising investment opportunities. After the boom has continued for some time, really profitable investment opportunities may be virtually "exhausted." Costs, which lagged substantially at first, tend to rise more and more rapidly as revival progresses. Raw material prices have risen sharply with the sharp increases in demand. Interest rates rise as banks and individuals find their supplies of "surplus" funds decreasing, and loanable funds become less easily and cheaply available. Eventually, excess reserves are exhausted, placing a ceiling on further increases of the credit so vital to increasing production and business expansion. Especially important also is the increase in wage rates, resulting from the increased demand for labor to carry on increased production. As long as large supplies of unemployed labor remain, entrepreneurs can hire additional workers without having to offer much higher wages, but once the supply of unemployed labor decreases substantially employers find themselves bidding against each other, bringing rising wages at an increasingly rapid rate. This increase is much hastened in industries where there are strong union organizations, and in such industries costs tend to catch up with revenues first, making further investment in those particular industries relatively less profitable than in other fields.

Some industries find certain types of highly skilled labor or very specialized machinery essential for their operation. As they increase output sharply during revival entrepreneurs find that finally they have pretty well exhausted the supply of such highly trained specialized agents. To get more laborers, for example, entrepreneurs must offer higher wages to induce longer or harder work or to bid workers away from employment in the factories of competitors. This occurs in many fairly small, but highly important, economic sectors as revival progresses, even though there may be still large supplies of unemployed resources fitted for other tasks. For example, in even the moderate revival of 1936 there developed a great shortage of skilled machinists and electrical workers, causing the wages of such labor to rise sharply. But since such critical agents are essential if production is to be increased further, entrepreneurs are forced either to pay the higher costs or to refrain from increasing production to meet the growing demand. Such shortages, which may be of skilled labor of particular types, specialized types of machinery, or whatnot, are usually called "bottlenecks," and act as very real checks upon further expansion wherever they occur.

In prosperity, as fewer and fewer unemployed remain, there is a further general increase in wage costs owing to the necessity of employing less and less efficient workers to supplement those retained during depression. Moreover, as production is speeded up and fixed plant utilized more fully, the efficiency of operation so carefully cultivated during the lean days of the depression may tend to disappear—partly because it is impossible to exercise such care and efficient management as production activity becomes more and more feverish, and partly because with increased profits less attention is paid to obtaining the utmost efficiency of operation at all times.

In addition to these rising operating costs, relatively fixed costs gradually rise. New long-term capital must be borrowed at higher rates; rents tend to rise, and although many are determined by long-term leases these may come up for renewal during this period; salaries, stickier than wages, are increased as revival continues; and depreciation and obsolescence charges also rise, especially as old machinery and methods are scrapped more readily to make way for "bigger and better" equipment intended to meet an expected steadily growing demand.

Revenues, on the other hand, which have risen rapidly during the earlier stages of revival, eventually tend to slacken their rate of increase, and the price situation can be characterized as shifting and uncertain. Thus rising costs become especially serious. The reason for this changed price situation is several-fold. The essence of revival is a sharp increase in the physical production of both consumers' and producers' goods, with capital goods most important. It takes a considerable amount of time to build, let us say, complicated new machinery intended to produce rubber tires, shoes, or whatnot. While the machinery is being built, wages are being paid out, and to finance the construction more credit is being put into circulation by the banks, and these factors increase the

demand for consumers' goods. However, the new consumers' goods are not yet being produced, so that prices tend to be bid up. Eventually the new machines begin producing rubber tires, shoes, and so on, and as a larger supply of these goods flows onto the market to meet the increased money demand the rise in prices is checked. Though the lag during which producers' goods are being made but no additional consumers' goods are being produced should not be overemphasized, it has in some instances evidently been of considerable importance. As revival approaches prosperity, the flow of consumers' and producers' goods onto the market increases rapidly owing to the heavy investment in productive equipment that has been taking place. But the demand for both types of goods tends to level off, and, in the case of producers' goods, even to decrease.

In considering this behavior of demand, let us examine first the demand for consumers' goods, distinguishing between the demand of consumers at retail and that of merchants at wholesale. The latter is likely to fluctuate more sharply than the former, because merchants hold stocks on hand, and the size of these stocks is varied considerably depending on the expectations of the merchants. If merchants expect demand to continue to increase they will lay in larger stocks, building up inventories; and if they expect prices of the goods at wholesale to rise they will buy at once to beat the price rise. Both these factors combine to explain the sharp increase in the demand by merchants for consumers' goods during the earlier phases of revival. But as revival proceeds, inventories are built up and merchants become hesitant about further increases in their stocks, partially because of lack of space and partially because of the expense of having funds tied up in inventories. Through revival the rate of increase in purchases is that of the increase in consumer demand plus the increase resulting from the building up of stocks on hand. Thus in prosperity, even though merchants expect good times to continue, there will be a slacking off of orders for consumers' goods owing to the fact that business inventories are no longer being increased.

The demand of actual consumers at retail ordinarily continues to increase through revival and into prosperity, but the rate of increase tends to slacken off as prosperity is reached. Consumers' replacement demands stored up during the depression have been pretty well filled and therefore this extraordinary source of demand tends to drop away. Moreover it is usually true that as the size of an individual's income increases, he will not spend the entire increase on consumption but instead will save part. This slacking off of the rate of increase in consumer buying of course makes merchants still less inclined to go on increasing inventories, and may even lead to a cessation of the increase in merchants' demands for consumers' goods, which, of course, is the demand that matters to the producers of goods. In addition to its effect on prices, therefore, the leveling off or decline in consumer and merchant demands for goods may lead under the acceleration principle to a very sharp falling off in the demand for the producers' goods behind those consumers' goods. In many cases even a decline in the rate of increase of con-

sumers' demand for a product may be sufficient to bring about a positive decrease in the demand for the producers' goods used in making it. Thus prosperity demand conditions may well combine with rising costs to check the upswing.

What of the funds that are saved out of increased consumer incomes in prosperity? These funds may be invested or they may simply be held, or hoarded. If invested, the money obviously goes to speed capital formation through construction activity, manufacture of machinery, and so on, for these are the types of projects in which savings are primarily invested. Or the saver may simply put his funds in the bank, not desiring to invest them. If the bank then makes loans and investments on the basis of the new deposits, obviously the result is about the same as if the saver had invested the funds himself, although in the second case they may go to different borrowers. The funds, if invested, tend to speed and intensify the activity in the producers' goods industries. But the banks do not necessarily increase their loans and investments. In many cases the money saved and deposited will just have been withdrawn from another bank account to pay the saver, and thus there is simply a transfer of accounts and no increase in reserves on which the banking system can expand loans and investments.

Many people fear that as this saving out of increased incomes progresses, so much will be invested in producers' goods that so many consumers' goods will finally be produced that they cannot be sold, especially since in prosperity people are likely to spend a smaller proportion of their incomes on consumption than in low income periods. This is largely a misconception. Unless the new productive investments promise to enable goods to be produced more efficiently and cheaply so that they can be sold at lower prices, thus permitting the larger supply to be disposed of, the investments will not appear profitable and will not be made. Actually, of course, this process does not work at all perfectly. There is much room for mistakes in judgment, especially since investors and producers have to calculate on the basis of expectations, not certainties. During prosperous periods expectations are likely to be considerably overoptimistic, leading to what may loosely be termed "overinvestment" in these industries—that is, a greater amount of investment in producers' goods than can be profitably used in making goods to sell at current (or expected) prices. But over the entire economy this difficulty by itself would not be serious. It is only when it is taken in conjunction with the cost increases, the demand decreases, and with other factors yet to be noted that it helps lay the basis for a severe decline in the level of economic activity.²

² On the other hand, if the funds saved by those receiving higher incomes are not invested, but instead are held idle, the result may be much more disastrous. For this means not simply a malallocation of resources, but a positive decline in the total amount of monetary demand. And it is upon the basis of an increasing, or at least stable, monetary demand that the entire revival and prosperity ultimately rests. Although hoarding is very unlikely to increase during prosperity, it becomes a major source of difficulty, once the decline gets under way.

An important "real" factor that underlies the upswing of revival and prosperity constitutes also one of the crucial causes of the downturn. This is the great "bunching-up" of investment in critical sectors of the economy during the boom. This effect can be clearly observed in the vast building boom of the 1920's. Not only were great numbers of houses erected, but a tremendous volume of office and industrial construction was undertaken and completed in the space of a few years. Under the impetus of inflated expectations, speculative building activity spread like wildfire. It has been estimated that enough skyscrapers were erected in three years to serve the country's needs for ten times that long. This building boom gave rise to far-reaching new employment and generated increased incomes in a multitude of related industries; indeed, the construction boom was at the center of the entire prosperity of the 'twenties. Yet by its very intensity the burst was inevitably short-lived. By 1927 large-scale building activity was largely spent, with only residential housing an attractive field for new construction—and building costs had risen so high as to make home building almost prohibitive. Perhaps overall booms would be impossible without such great bursts of activity in durable producers' and consumers' goods industries, but general booms based largely on such bursts can rely on this impetus only for a limited period. And the end of a concentrated spurt of activity can seldom be a leveling off at the peak, unless the good produced is very nondurable. In most cyclically critical industries, if we have the spurt of activity we must also have the inactivity that follows as the economy digests the vast output bunched into the boom period.

A brief digression is worth while on the revival and prosperity that occurred in this country during the late 1920's, since it differed in one major respect from the usual pattern. During this period, the wholesale price level remained almost stable; yet a boom of fair proportions in business and industry and intense speculative activity in securities developed. Owing primarily to the extraordinarily rapid rate of technological advance, costs declined steadily, falling away from the roughly stable prices, which were held up in part by large-scale creation of new bank credit instead of falling apace with costs. Under this "profit inflation" business and industry were stimulated to increased activity. There occurred the same boom psychology, the same heavy investment in capital goods industries, the same increased buying and expansion of bank credit, as in other booms. Since then, organized labor has become much stronger; and it seems likely that under similar future conditions wage rates would probably be pushed up more rapidly, thus absorbing a large share of the cost-savings from technical progress. The real danger in future recovery periods seems to be more that wage rates will be pushed up so rapidly (at least in many crucial industries) that re-employment will be checked before prosperity is reached.

For the many reasons indicated, as prosperity proceeds the economic system becomes more and more "vulnerable" to shocks tending to precipitate a downturn. What actually brings on the turn (which may or

may not take the form of a financial "crisis" in its earliest stages) varies from cycle to cycle. The basis for the high level of expectations in the business and financial community, upon which the boom rests, becomes considerably less solid as prosperity proceeds. Hence, a fairly severe shock, such as the failure of an important business firm or bank, is increasingly likely to prick these inflated forecasts. In a majority of turning points, especially in those that have been accompanied by financial panics, some such factor as this has helped set off the process culminating in crisis and temporary panic, and followed by deepening depression. But this has not been true of all cases, and it is much more important to recognize the various factors making the downturn more and more likely than to speculate on what particular factor tips the balance in particular cases. Rising costs, exhaustion of excess bank reserves, the piling up of inventories, the slackening off of demand for durable consumers' and producers' goods under the principle of derived demand, the growing scarcity of large-scale profitable investment opportunities—in the setting provided by such basic self-limiting factors, any one of many events may cause the downturn to occur just when it does. Any explanation that blames the downturn completely on one factor is at least 99.99 per cent sure to be wrong. Some one particular event may be the spark that sets off the fuse, but in most cases it is highly uncertain what even this event is.

Recession

Once something has happened to destroy the aura of success surrounding business and industry, the path toward depression is a direct and cumulative one. Businessmen, somewhat disappointed in sales and profits as prosperity levels off, are uneasy. Under such conditions any sharp disappointment of expectations or widespread shock may exercise a very strong psychological effect. Sharp curtailment of investment and production programs may quickly result. Since the level of productive activity and investment is determined primarily by the relations of *expected* selling prices and costs, an expected decrease in demand has precisely the same effect as would an immediate decrease that was expected to continue, so far as the effect on production policy is concerned. Indeed, an expected demand decline may be even worse than an existing lower but stable demand, since businesses may unload inventories on the market in an effort to sell before demand conditions worsen. This action leads to the very price declines sellers are trying to "beat" and to still further worsened expectations for the future. This breakdown of expectations is especially severe whenever costs have risen substantially during the revival and prosperity, so that even a minor decline in expected selling prices is likely to wipe out prospective profit margins. Even though output and cost conditions remain unchanged, it no longer appears profitable to maintain output; workers are laid off—and the army of the unemployed begins to grow.

No matter what the factor immediately precipitating the downturn

may be, there is one factor that is the *necessary* concomitant of every downturn that develops into a severe depression—a contraction of total monetary demand. Unless the rate of spending decreases and the banks exert deflationary pressure by calling in loans and liquidating investments, it is inconceivable that recessions could reach anything like their past severity of decline. The deflationary pressure of hoarding and credit contraction has been the core of virtually every serious cumulative recession.

The reason why monetary contraction is the necessary core of cumulative deflation should be apparent in view of the central part played by monetary demand in the cycle. If monetary demand were maintained roughly constant following the downturn, there might still be something of a recession during which overly large inventories would be unloaded, the super-era prosperity plans abandoned, “overinvestment” in particular industries eliminated, and so on. But if total monetary demand were not to decrease, there could be no very substantial decline in prices or in sales volume, and before long businessmen and investors would recover from the unfavorable expectations involved in the downturn. Events would turn out to be not so bad as had been expected, and expectations would be revised upward.

However, once recession has set in, total monetary demand is likely to contract cumulatively. Bankers, as well as businessmen, have expanded to the “limit” during the boom times, and their expectations, like those of businessmen, are likely to shift sharply downward. But once bankers become pessimistic about the future, their power for contraction far exceeds that of ordinary businesses. The first, and entirely natural, reaction of the banker is to strengthen his own extended position by calling in loans that seemed entirely sound on the basis of the optimistic expectations of the past but which are of very doubtful safety in view of changed expectations. Until the insurance of bank deposits by the federal government in 1933 (through the Federal Deposit Insurance Corporation), crises usually produced depositors’ “runs” on banks—mass withdrawals of deposits for fear the banks would close and be unable to pay off depositors. Such runs placed on the banks tremendous pressure to liquidate assets. Since the inauguration of the FDIC this particular danger has largely vanished save in very severe crises. However, since not all bank deposits are guaranteed (only up to \$5,000 for each depositor) it remains conceivable that serious mass withdrawals by uninsured depositors could occur.

Whether it be in order to get cash to meet large withdrawals or simply to protect against possible loss on loans and investments, banks move to liquidate loans and investments once recession begins. But in such a period securities can be sold only on an already demoralized security market, and calling loans forces business enterprises to one of two alternatives: (1) the business enterprise can take special measures to get cash, as by forced selling of its inventories, which acts to depress prices and reinforces the deflationary pressure; or (2) it can obtain the funds to

repay by borrowing from other banks or by selling securities which it owns. But at such times borrowing from other banks is not easy, and if it could be done the repayment of the loan to bank A would constitute a drain on the lending bank of roughly the amount repaid to bank A. Moreover, it is no easier for private businesses than for the banks to sell securities in such periods. Business and industry in general simply cannot liquidate their indebtedness to the banks on short notice. Borrowing is often theoretically for short periods, but renewal practices are so commonly counted on and accepted that in essence business and industry are borrowing from the banking system at long term, often primarily for fixed and semifixed capital. Thus, when the banks exert severe pressure to liquidate business loans on very short notice, the result is likely to be complete demoralization. In order to get funds, businesses are forced to dump goods and securities on the markets for what they will bring, and with large numbers doing this at once prices decline drastically. The further prices are depressed, the more difficult it is to obtain enough funds to repay the bank loans, and the more necessary it becomes for bankers to call loans to protect their interests.

It must be remembered, moreover, that if currency hoarding occurs the mass contraction by the banking system is not simply a dollar-for-dollar calling of loans to pay out cash. On the contrary, if the system is fairly well loaned up to the limit at the peak of prosperity every dollar withdrawal in cash forces contraction of roughly five dollars in loans and investments, if an average 20 per cent reserve requirement is assumed. Even though a banker dislikes to put pressure on his borrowers, he is forced to do so by the drain on his reserves from adverse clearing balances as other banks liquidate loans and investments. The monetary forces at work could not be more completely perverse had they been diabolically designed for the purpose. Not only does the rate of spending drop sharply with growing uncertainty and pessimism, but at the same time a very sizable proportion of the entire circulating medium is actually destroyed as the banks are forced to contract loans and investments.

Forced selling of business inventories and securities is especially serious because on the demand side forces are also at work to reinforce the decline. Individuals' demands for many types of goods tend to level off in prosperity, and with the uncertainty and loss of confidence involved in a deflationary recession, people are certain to spend less and hoard more, waiting to see what the future will bring. Simultaneously, growing unemployment means that individuals are not only likely to hoard more but will have less income from which to spend and hoard, resulting in a "double decrease" in the velocity of circulation of funds. Feeding on itself, increased hoarding chokes off production and employment.

The general nature of the cumulative deflationary spiral should now be clear. Declining prices are followed by expectations of further declines, cutting down the profitability of operations—lower production with resulting unemployment and lowered money incomes—further hoard-

ing and pessimism, leading to still further declines in prices and output—cessation of new investment and construction, the decline speeded by the downward action of the acceleration principle as consumer demand falls off—and forced liquidation under pressure of the contracting banking system, which is itself involved in the spiral. Once under way, contraction and deflation develop cumulatively into full-fledged depression.

The part played by sticky costs in the downward spiral has been emphasized by many economists. If costs were not sticky, falling selling prices would not mean sharply decreased profits, and the seriousness of the contraction of monetary demand might be greatly mitigated. Among operating costs wage rates for strongly unionized groups and prices of monopolized materials and parts are especially sticky downward. Also, during prosperity many firms build expensive plants and equipment, upon which depreciation, obsolescence, rent, and taxes become very heavy burdens in recession. Railroad and public utility rates are likewise extremely sticky downward in most cases, and such rates are important costs to many types of industries. Without contracting monetary demand sticky costs would not be so serious, but sticky costs in the face of falling selling prices lead to contracted production and investment and to unemployment.

Lastly, in spite of their prominent position in the public eye the relation of the security markets to the cyclical process is generally a direct but not a crucial one. Security (that is, stock) prices have almost invariably preceded cyclical movements in general business, though by no definite lead. As prosperity increases and boom expectations become more and more prevalent, speculative activity almost always becomes very great on the stock exchange. Especially was this true in 1928–29, when such speculative activity reached heights completely out of touch with the actual profit possibilities or the general level of business activity. Such a speculative boom carried within itself the causes of its own undoing. Once the bubble was pricked, as it was in late 1929, there was a precipitate decline in security prices, reinforced by calls of the New York bankers for repayment of loans made to finance margin purchases. The result was such a mass destruction of paper profits as the world has never seen before. This crash is significant for cycle purposes largely because of the psychological reaction to this spectacular end to the “new era” in the stock market, although the direct repercussions were great on the issuing firms and on people whose fortunes were wiped out. The break in the stock market provided the psychological turning point of the boom—people found their airy dreams ruined, their financial security vanished, their certainty of permanent prosperity and an easy future turned to utter uncertainty. A stock market crash was, and has often been, the signal that appears to have set off the downward spiral, although the problem of what set off the stock market crash just when it occurred remains unanswered. Financial crisis has usually but not invariably characterized the upper turning point of the cycle. Yet, probably in most cases the significance of such crises has been as a “signal”

for the downturn, not as an inherently necessary step in the cyclical process.

Conclusion

The following simple outline lists some major characteristics of each phase of the cycle in order to summarize briefly some of the complex interrelations which have been under discussion:

Depression—

- Low level of income
- Widespread unemployment
- Nonreplacement of worn out capital goods—little new investment
- High rate of bankruptcy—inefficient firms being eliminated
- Costs falling to low levels—catching up with prices
- Low inventories
- Postponement of purchases and accumulating need for durable goods
- Excess bank reserves
- Hoarding
- Development of profit possibilities not yet exploited
- Uncertainty and pessimism prevalent.

Revival—

- More favorable expectations
- Buying to replace worn out durable consumers' and producers' goods
- Increasing investment (in new and old industries)
- Increasing employment
- Increasing money incomes
- Rising prices—usually faster than costs
- Increasing inventories, to "beat price rises"
- Increasing bank loans and investments
- Dishoarding—increased velocity of circulation
- Expectation of further price rises
- Operation of the "acceleration principle" upward.

Prosperity—

- Continuation of favorable expectations
- Replacement of worn out durable goods
- High level of output—increased production of new capital goods
- Relatively full employment of productive agents
- Costs catching up with revenues
- Inventories large
- Speculative activity in many lines
- "Exhaustion" of profitable investment opportunities
- End of "acceleration principle" upward
- Excess bank reserves depleted—credit expansion checked
- Failure of speculative expectations to materialize
- Leveling off of profit margins—beginning of declines.

Recession—

- Yielding of favorable expectations to uncertainty and pessimism
- Declining profit margins
- "Acceleration principle" working downward
- Possibly some "bubble-bursting" crisis
- Decreasing investment—unemployment of productive agents
- Decreasing money incomes

Falling prices and costs—costs lagging

Liquidation of securities and inventories at falling prices—selling to “beat the price decline”

Contraction of bank credit

Increased hoarding

Falling levels of output.

While there are several factors (especially the exhaustion of banks' excess reserves and the exploitation of major investment outlets) that are almost certain to put a “ceiling” to the boom, there is much less certainty as to the “floor” in the depression. Why the bottom is reached just when and where it is is most difficult to explain. One partial explanation advanced is that after prices have fallen so far and after much liquidation has occurred, people begin to look for a leveling off, possibly on the basis of past experiences. Further, buying that has been postponed throughout the decline tends eventually to become more “necessary,” especially when expectation of further price declines weakens—an overcoat may be used a third or fourth year but eventually it becomes almost unusable. Such replacement demand, by individuals and industry, may amount to considerable proportions in comparison with depression lows. Another factor sometimes stressed is the increased business efficiency that comes with hard times. Forced liquidation is eventually completed, and this deflationary pressure removed. Currently many writers feel that the development of new products and methods providing profitable investment outlets is perhaps the most important revival stimulus. And of general but crucial significance is a lessening of the pessimistic, uncertain expectations of entrepreneurs and investors. Often this shift is dependent upon external factors—the most striking example is the great revival of confidence following the bank holiday and positive government action in March, 1933. To explain revival each depression must be considered as an individual case. As with the upper turning point, the underlying “favorable” conditions for the change are usually more important than any one particular event that may shift the balance. Only very recently have intentional government corrective efforts come to play a major role in the recovery picture.

One important aspect completely omitted from discussion thus far is the international nature of business fluctuations. Over the last century there has been a marked similarity of business fluctuations in the major industrial countries. This is due to the close economic and political interrelations between nations, especially under the international gold standard of past years. This does not mean that cycles in various countries have been identical—rather, simply that cycles have generally tended to be international rather than confined to one country. This aspect of cyclical fluctuations is omitted here since international economic relations are yet to be discussed; it will be considered in that setting.

Further, no mention has been made, except incidentally, of agriculture's place in the cyclical process. Although good or bad crops do have some influence on business fluctuations, apparently cycles of good and bad

times in agriculture are largely a result of prosperity and depression in the business and industrial world. The demand for farm products depends largely on business and industrial conditions. In prosperity industry takes large amounts of farm products as raw materials; and wage earners, fully employed at good wages, buy large amounts of foodstuffs. In depression, industrial demand for raw materials falls off; and wage earners, unemployed or employed at low wages, are forced to contract consumption of foodstuffs and of finished products that use farm products as raw materials. Although there is a close interrelation between agriculture and business and industry, the causal connection is apparently more from business fluctuations to agriculture than vice versa.

It may be useful to summarize the complex cyclical process briefly: Excluding government activity, the level of employment and production (investment) is determined by entrepreneurs' profit expectations. These expectations depend on the entire complex of economic, social, and political conditions at any given time, but essentially they reduce to expectations as to cost-revenue relations over the period for which investment and production plans are made. Fluctuations in investment activity and in monetary demand for producers' and consumers' goods tend to become cumulative. But many costs are sticky, moving less quickly. This fact gives rise to widening profit margins on the upswing, and decreasing or vanishing ones on the downswing. This combination, coupled with the fact that most entrepreneurs are essentially optimistic once recovery is under way so that they project better conditions into the future and correspondingly pessimistic in deflation, explains the upswing of revival and the downswing of recession. A movement in either direction tends to feed on itself cumulatively. Yet in the upswing as speculative profit expectations are finally disappointed and "profitable" investment opportunities become scarcer, prosperity tends to reach its peak and level off or fall. In the downswing as conditions prove to be not quite so bad as expected when liquidation is completed and falling costs "catch up," depression tends to reach a bottom and level off or rise. On this simple pattern practically all the factors at work in business fluctuations can be fitted—one should be careful not to lose sight of the woods for the trees and fall into the error of supposing that any one or a few particular happenings "explain" cyclical fluctuations. All factors are important as they influence individual and business expectations—ultimately as they influence the relation between expected costs and revenues.

CHAPTER 45

Economic Policy and Business Fluctuations

THE great severity of the depression following 1929, coupled with the fact that by 1940 there were still around nine million unemployed in this country in spite of a return of industrial activity to high levels, caused many persons to forget an important historical fact. For well over a hundred years the United States has been experiencing alternating waves of good and bad times, and several of the depressions have been both long and severe. Never before the 1930's had there been any serious governmental effort to "cure" the depression—rather, events had been left to run their course and "cure" themselves. And in spite of the fact that no important remedial steps were taken, we had revived from each depression period to go on to a succeeding prosperity. This same general experience of revival without remedial effort had been true of the other industrial countries that had been experiencing business cycles.

In the past few years people have become so accustomed to thinking of revival in terms of government spending and other remedial measures that it is easy to overemphasize that aspect. Indeed, it is not yet entirely clear whether the conglomeration of supposedly remedial measures undertaken by the federal government after 1930 did more to help or to hinder revival. In spite of the remedial efforts, this last depression was the most severe and one of the longest in history. Only with the "war" boom of 1941, based upon a tremendous program of spending for rearmament, did many indexes of business activity and employment surpass the peak 1929 levels and unemployment fall under five million. Why did revivals come without government remedial effort all through our history, and only so slowly in spite of the assistance of government "remedies" during the last decade? What can be said regarding the usefulness of various government policies intended to stimulate recovery and to maintain prosperity once attained?

The preceding chapter has described in some detail the numerous self-perpetuating tendencies inherent in cyclical business fluctuations, once these fluctuations are well under way. It is relatively clear why a boom or recession, once well started, moves cumulatively upward or downward. But it is much less clear exactly why this cumulative movement reverses itself—why a boom turns into a recession, and vice versa—although various factors exist that are likely to bring about these turning points. Perhaps the greatest problem is why depression has always been followed

by revival, usually at rather regular intervals, without any formal remedial action.

There can be little doubt that the resilience of the economic system in springing back recurrently from depression lows over the last century and a half has been due in considerable part to the rapid expansion of the economy and to extraordinary "external" stimuli that came at propitious moments. The very rapid geographical expansion of our economy with the opening of the West and the correspondingly rapid growth in population provided an expanding range of new investment opportunities, and provided a momentum that did much to lift the country out of depression periods. At the same time, the amazingly rapid technological advances of the nineteenth and early twentieth centuries opened vast new fields for profitable investment, especially in such heavy investment industries as canals, railroads, automobiles, electricity, and so on. Likewise, during much of the nineteenth century international economic intercourse was relatively free, providing both a heavy inflow of foreign capital to finance rapid economic expansion and a market for many export products.

In addition to these underlying expansionary influences, on several occasions "external" events were of crucial importance in expediting revival. The great gold rush to California in the late 1840's was one such event. The Civil War, with its large expenditures, was another. Inflationary results of the Sherman Silver-Purchase Act played a major part in the recovery from 1890 to 1893. The Spanish-American War and the great gold discoveries in Alaska and South Africa provided the impetus for the recovery of 1897-99. Thus, the self-perpetuating forces in business fluctuations may have played only a minor role in explaining the revivals following each depression period.

During the long depression of the 1930's there was a widespread wave of pessimism, a feeling that the economy had finally reached a state of "maturity" and that the expansive forces of the nineteenth century could no longer be counted on to help bring revival by opening new profitable investment opportunities. Some people went so far as to argue that the economy was now "stagnant," pointing to the long and severe depression in spite of huge government expenditures and other policies aimed at bringing revival. The important characteristics of this "mature" economy were asserted to be:

1. Decline in the rate of growth of population, which cuts down the rate of growth of demand for houses, clothing, foodstuffs, etc.
2. End of geographical expansion—closing of the geographical frontier for capital investment and expansion.
3. Increasing efficiency of capital due to technological advance, so that workers find it increasingly hard to be reabsorbed after being displaced by machines.
4. The continued accumulation of capital, which tends to reduce its yield as opportunities for new investment are used up.
5. The trend toward economic nationalism, which discourages foreign trade and loans.

There is no doubt as to the facts—all of these tendencies do exist. And they all may be of importance. But it seems likely that many writers greatly overemphasized them and hence took an overly pessimistic view in saying that private investment and business could not be expected permanently to reabsorb the unemployed.¹ To the extent that these tendencies exist they made revival during the 1930's less likely than it would have been a century before. It surely is true that no major technological advance happened along to provide an opening for large-scale investment at the needed moment, for example, as did railroad building several times during the nineteenth century and as did the auto industry and construction following the depression of 1920-21. But to argue from this that we cannot expect more technological advances, which will open avenues for widespread investment, appears very questionable. Less pessimistic writers argue instead that the failure of private industry to absorb the unemployed during the 1930's was primarily explainable on several other grounds, and that, even though the "maturity" argument has some validity, there is strong reason to believe that private investment (business) would have revived much sooner if conditions had been more favorable and certain barriers had been removed. With the rearmament boom, talk of "maturity" has largely vanished in the face of the highest levels of productive activity and employment yet known, far surpassing 1929 levels. Yet, there can be no escaping the fact that the boom has rested very largely on the huge government rearmament expenditures, and to think that such a war boom permanently solves all unemployment troubles would indeed be short-sighted.

Most discussion of antidepression policy has been in short-run terms. We have typically worried about how to get out of depressions after getting in them. Comparatively little attention has been given to the more fundamental problem of long-run policies calculated to maintain economic stability and thus prevent the excesses both of the inflationary booms and of depressions. In this chapter some of the more important measures used recently in the attempt to control business fluctuations will be considered first. Later, in view of the shortcomings of the various current policy proposals, attention can be given to certain lines of long-run economic policy calculated to maintain economic stability, rather than simply to restore "prosperity" after depression.

The Basis for Economic Policies to Mitigate Business Fluctuations

If revival came more or less automatically from all the depressions of the last century, why is it desirable to undertake any governmental policies to achieve this end? The answer is, few human trials are as great as those forced upon great masses by the unemployment and poverty

¹ Similar pessimism, and even use of this same concept of a "mature" economy, can be found in several of the earlier severe depressions, especially the long depression of the 1870's.

of depression periods, and it may be possible to bring revival more quickly and surely by a wise governmental antidepression policy than would occur "automatically." Especially is the case for positive remedial action strong if the arguments as to the "maturity" of the economy are valid enough to mean that revivals from "natural" causes will henceforth be very slow and uncertain.

Severe business fluctuations depend in considerable part on fluctuations in money income and demand, especially in view of the many "sticky" costs that mean fluctuations in profits as selling prices move up and down more rapidly than many costs. The whole process centers around the profit expectations of entrepreneurs and investors, which are based not only on existing profits (price-cost relations) but more on expectations of future changes in revenues (money demand conditions) and costs. A vast range of considerations, many "noneconomic" in nature, affect profit expectations at any time and all these must be considered in formulating economic policy designed to mitigate business fluctuations.

Since monetary demand plays such a central part in business fluctuations and since it is the factor perhaps most directly amenable to control, major attention has been centered on this approach in business cycle policy. Considerable attention has, however, also been paid to the possibility of controlling costs and individual price-cost relations. Besides these two basic avenues of approach, numerous other suggestions have been offered. For purposes of convenience, the various business cycle policies are considered under four headings: (1) monetary policy; (2) fiscal policy (government finance); (3) policy toward costs; and (4) miscellaneous policies. Actually these various types of policy are very closely interrelated, as will be apparent.

Monetary Policy

Monetary policy includes control of money and credit. Monetary policy and fiscal policy (public finance through spending, taxation, borrowing, and so on) are very closely interrelated and properly should not be separated—they are segregated here primarily for convenience in discussion. Acting directly and with the Federal Reserve System, the government can attempt to check undesirable booms by withdrawing funds from circulation through taxation and by control of bank credit in various ways; it can attempt to stimulate revival from recession or depression by putting newly created money into circulation, by attempting to increase the velocity of existing money and credit through revenue and expenditure policies, and by stimulating the expansion of bank credit.

Issue and withdrawal of money

The simplest way in which the federal government (or the Federal Reserve) might use its monetary powers to mitigate business fluctuations would be simply to issue and spend new money when it wanted to stimulate revival, and to withdraw (by taxation) and destroy money whenever it wanted to check an undesirable boom. By so controlling the amount

of money in circulation the government would be directly controlling fluctuations in the money demand for goods and services, and thereby influencing business conditions.

Against this *prima facie* case for the most direct method of monetary stabilization, certain criticisms are raised. Aside from the very difficult problem of when to issue or withdraw money so as to promote business stability at the desired level, there are said to be two primary weaknesses to this plan of action. The first is that actual money is a small part of the total means of payment—changes in the amount of bank credit are quantitatively much more important. Withdrawal of money by the government to offset a boom might easily be more than offset by changes in velocity and by new bank loans and investments made in the optimism of the boom period. A similar difficulty would arise in so trying to check a recession. Large issues of money in depression over the amount wanted for use and hoarding by the public would ordinarily find their way back into the banks where the cash might be used as the basis for an increase in credit; however, this is likely to be of limited use in such periods when there are large excess reserves anyway. The second weakness is that many persons (especially conservative businessmen and bankers) have an ingrained fear of “inflation” from the issue of new money. Often this fear is partly irrational—based on the assumption that an increase in the quantity of money will necessarily lead to a runaway price inflation and eventual bankruptcy of the government. But whether the fear is irrational or not, the issuance of large amounts of new money to check recession may do as much harm as good, by creating uncertainty and pessimism in the minds of businessmen and investors, making them all the more hesitant to expand output and undertake new investments. Ordinarily there seems to be much less fear of “inflation” through expansion of bank credit, although actually expansion of bank credit has the same effect on prices as new money and is likely to be quantitatively much more important. A related suggestion, considered later, is that the Federal Reserve create bank credit directly rather than actual money.

Treasury control of bank credit

The primary means of monetary control has been and promises to continue to be through control of the expansion and contraction of bank credit. Both the federal Treasury and the Federal Reserve have important means of control over the quantity of bank credit, though neither nor both together have anything like complete control. The Treasury exercises its power over bank credit primarily through its financing policies. If it wants bank credit expanded (for example, to finance a depression spending policy), it may achieve this result by issuing new bonds and selling them to the banks.² When the bank buys bonds, it

² This is simply borrowing from the banks. It should be kept distinct from the sale of already existing bonds by the Federal Reserve in order to draw down bank reserves. In the Federal Reserve case the bank immediately loses reserves to the extent of purchase without losing any deposits—so excess reserves are reduced and

expands credit just as if it had made a loan to an individual or business, and the Treasury obtains its funds through an expansion of bank credit instead of drawing on funds that individuals might otherwise have spent themselves. The Treasury may also stimulate credit expansion by the issue of new money that finds its way into the banks as new reserves, but the short-comings of this method have already been noted. Ordinarily the Treasury has been interested in bank credit primarily as a means of obtaining desired funds to spend in depression or war periods. It has made little effort to check credit expansion in boom periods, since expanding monetary demand can be checked more directly through new taxes if the Treasury wants to take action. Control of credit expansion has usually been considered a function of the Federal Reserve. But if the Treasury has large deposits with the commercial banks, it can draw down excess reserves by transferring these deposits to its accounts at the Federal Reserve Banks. This transfer means adverse clearing balances for the member banks losing government funds, just as when an ordinary customer withdraws his account by check and transfers it elsewhere. Since the Treasury deals in such huge sums of taxed and borrowed funds each year, its power over the quantity of bank credit is tremendous. The astronomical expansion of bank deposits since the outbreak of war demonstrates the overwhelming position of such government policies.

Federal Reserve control of bank credit

The Federal Reserve is commonly thought of as the agency having the power to control bank credit and this has always been one of its major purposes. Although the Federal Reserve authorities have been given direct responsibility for credit control, actually their powers are substantially less than these potentially open to the Treasury through its fiscal policies. The Federal Reserve has several methods it may use in attempting to check an undesirable credit expansion: (1) making open market sales, (2) raising the rediscount rate and the rate on advances, (3) direct pressure on banks to stop expansion, (4) raising legal reserve requirements, and (5) tightening credit in the security markets and in installment buying. The converse steps may be used to stimulate revival. These methods have been described already. Here it is sufficient to note again briefly how effective they are likely to be.

Federal Reserve authorities may be able to check credit expansion in the upswing if their resources are not exhausted before the excess reserves of the member banks are used up. By open market sales the Reserve may draw down excess reserves and hence limit expansion, but this is possible only as long as it has securities to sell in larger volume than existing excess reserves. Raising the rediscount rate or the rate on advances is effective in checking expansion only if member banks have to borrow, and as long as they have excess reserves borrowing is unnecessary. The

pressure is exerted to contract credit. In the Treasury case, the Treasury takes the bond proceeds in the form of a deposit at the bond-buying bank, and credit is created, just as if the bank had loaned to an individual or business concern.

ability of the Federal Reserve to check credit expansion ultimately depends on its ability to exhaust member banks' reserves. The 1941 boom period clearly illustrated this limitation. The Board of Governors had power to raise requirements to 26 per cent, 20 per cent and 14 per cent for central reserve city, reserve city, and country banks, respectively. This increase, coupled with open market sales of securities held, would have used over half the existing excess reserves, but there still would have remained very large excess reserves even though the Federal Reserve's resources were exhausted. Federal Reserve power to limit bank credit expansion, though substantial, was obviously inadequate. However, ordinarily the Federal Reserve authorities have had considerable power to check credit expansion, if they were willing to take sufficiently drastic action. Recently the obligation that the Federal Reserve authorities have felt to assure easy financing conditions to the Treasury in the war period has largely eliminated the possibility of vigorous anti-inflationary action.

Federal Reserve powers to check a contraction of credit or to stimulate expansion once the economy is in a depression are much more limited. In depression banks typically have large excess reserves. The Reserve authorities may lower rediscount rates and rates on advances, but if the member banks have no need to borrow, the lower rates have no effect. The Reserve Banks may buy government securities in the open market, thus adding to the excess reserves of the banking system, but the failure of banks to make loans and investments in depression is not due to lack of reserves. It is due, rather, to the unwillingness of bankers to make loans to any except "safe" risks and to the fact that such "safe" risks do not want to borrow—a major reason that they are considered "safe" is their strong cash positions. Although larger excess reserves give an added stimulus to the banker to expand credit, they are by no means sure to get results. Reserve requirements may be lowered, thus further increasing excess reserves—but again there is no surety that this will lead to credit expansion. But one vitally important power that they do have, and unfortunately have not always exercised, is the power to see that the banking system is never driven to mass liquidation by reserve shortages; this is perhaps the most vital of all Reserve functions.

Thus, the Treasury and Federal Reserve do have important monetary powers to mitigate business fluctuations, directly through the power to issue and withdraw money and indirectly through control of the volume of bank credit. Their power to check upward expansion is considerably greater than their power to stimulate recovery, especially in the case of bank credit control. In any case, control over the amount of money and credit in existence means far less than complete control over the price level and the level of business activity. The government has no direct control over the velocity of circulation—over the confidence of the people and their expectations—and often nonmonetary factors are more important than monetary ones in determining the level of production, employment, and income.

The problem of "timing" monetary policy

These limitations on the *powers* of the Treasury and Federal Reserve to stimulate revival and to check inflationary booms are only part of the problem. Almost equally vital is the question of when the powers are to be exercised. If business fluctuations were simple and precise, so that turning points were always sharply marked and even predictable, monetary policy would be relatively simple. But there is often a good deal of uncertainty as to just what sort of monetary policy should be followed, even assuming the ability of the policy to have the desired effect on business conditions. In 1929 many persons were arguing that credit should be expanded further lest the "new era" be checked; others demanded repressive action. In 1937 the situation was very mixed, with a high level of industrial production and with other favorable business indicators but with many millions still unemployed. Should restrictive measures be applied in such a case to safeguard against a runaway boom, or should credit expansion be encouraged? If stimuli to revival are not discontinued soon enough after "prosperity" has returned they are likely to lead to overexpansion and inflationary speculation. On the other hand, if the stimuli are discontinued too soon, the revival may die off and the system may fall back into depression. Although many indicators are now available as guides to monetary policy, the problem of timing remains one of the most difficult faced in attempting to achieve business stability at relatively full employment.

Federal Deposit Insurance Corporation

Crises and recessions have always been marked by "runs" on banks and waves of bank failures. As part of the emergency legislation of 1933 the federal government set up a plan of bank deposit insurance—the "Federal Deposit Insurance Corporation." The plan now guarantees that if any insured bank fails, the FDIC will pay off each depositor up to a limit of \$5,000; deposits above this sum are uninsured. The banks insured pay into the fund a certain assessment each year to build up a reserve to pay losses, just as in ordinary insurance. But what gives the plan its strength is the fact that the federal government is behind it—the assurance that the federal government stands ready to make good deposits in closed banks. Any bank may obtain insurance by meeting certain requirements as to soundness; and virtually all banks are now members. A large majority of all depositors are fully insured, though less than half the country's deposits by volume are covered because of limitation of the insurance to \$5,000 per account. Once depositors are confident that there is no danger of banks being unable to pay their deposits, the reason for bank runs vanishes. Only depositors holding partially nonguaranteed accounts any longer have any cause to institute bank runs.

By largely eliminating bank runs the FDIC has been of great aid in preventing financial crises, and should continue to be. Its importance is much more in preventing bank failures than in paying off the deposits

once banks have closed. In the past, mass withdrawals from banks at crisis periods have forced them to liquidate securities and call in loans with disastrous results for borrowers. Bank failures have played a central role in destroying public confidence in financial and business institutions. Since the FDIC largely eliminates bank runs, it largely removes the pressure on the banks to get cash suddenly and hence makes it unnecessary for banks to contract credit so rapidly once recession begins. The danger of financial crisis is much less than it has been in the past, and future recessions should be less severe than past ones, *other things being equal*, because the pressure on banks to liquidate has been alleviated.

Fiscal Policy (Government Finance)

The powers of the federal government to influence the level of production and employment have already been noted briefly in connection with "monetary" policy. But most of these powers of the government lie in the realm of what is usually called "fiscal" policy, or government finance. Roughly defined, fiscal policy includes taxation and borrowing as means of obtaining funds and government expenditures of these funds. It will become increasingly apparent that monetary and fiscal policy considerations ought not realistically to be separated—they are segregated here only for simplicity of exposition.

Fiscal powers to check inflation

Suppose a runaway boom is in progress, with speculative activity rampant and increased monetary demand resulting primarily in higher prices rather than increased production. What are the fiscal powers of the government to check this situation, in addition to the powers already noted under monetary policy?

In principle the answer is simple. The government can withdraw funds from circulation, presumably by taxation,³ and fail to spend the funds. This directly reduces the total monetary demand for goods and services and acts as a direct brake on the boom and price inflation. How severe the taxes would have to be in order to check the inflationary boom would depend upon the particular case in question. Almost any type of tax would obtain the desired effect of reducing the total incomes people have to spend and so checking the inflationary boom. Most direct would be a general income tax, taking a portion of every income before it could be spent. The choice between taxes would have to be made on several

³ Funds might be withdrawn by borrowing and then held idle, but this would probably be much less effective. Borrowed funds might well come simply from bank credit expansion or from sources where the money would have lain idle anyway. In these cases the action of the government would have little or no effect in decreasing effective monetary demand. Only if the borrowed funds were drawn from sources where they otherwise would have been spent would there be any decrease in effective monetary demand. By contrast, taxation is much more likely to withdraw active funds.

grounds, perhaps the most important being that of an "equitable" distribution of the tax burden. Comparison of the nature and effects of various taxes is included in later chapters;⁴ here it is only necessary to see the part taxation could play in checking a boom and/or price inflation.

In addition to the problem of proper timing for such taxes to check boom period inflation, a very important political difficulty exists. Boom periods are typically periods of unrestrained optimism—the "new era will last forever"; "things are going up and up." In the face of such a public opinion, legislators would be acutely unpopular were they to impose heavy taxes to check the boom. And quite aside from the political pressures involved, there is considerable doubt whether the congressmen could be expected to realize the need for taxation of this sort—congressmen are subject to the same "new era optimism" as are others. Even under obviously inflationary conditions, Congress is exceedingly hesitant to vote new taxes. Taxation to check booms and rising prices has almost invariably been "too little and too late," and there probably is little reason to expect any other policy in the future, *unless the policy of checking the boom is firmly adopted before the need for its application*. In a period of approaching boom and price inflation it is fairly easy to reflect coolly on the coming dangers and possibly even to enact a law providing for automatic imposition of taxes to check the inflation should the need arise. But to hope for the adoption of such a "sensible" policy during the boom period appears, in the light of history, to be a pious hope indeed. The events of the present war inflation period bear sad witness to this fact.

Fiscal powers to stimulate revival

The powers of "monetary" policy to stimulate revival from depressions have serious limitations, and, for the first time during the past decade primary reliance has been placed on "fiscal" policy, first as an unplanned result of relief spending and later as a conscious expansion policy. The primary method used has been the "unbalanced budget," the federal government greatly increasing its expenditures without a corresponding increase in tax revenues and the difference (or "deficit") being financed by borrowing. The rationale of the plan is simple. By borrowing idle funds and putting them to use, employment and production are increased, and national income rises. The borrowed funds can then be repaid out of taxes during the ensuing prosperity. Everyone would agree that it would be foolish to try to have receipts just equal expenditures every day, or even every week. Would it be foolish to take a month as the period for balancing the budget? Surely so. Why then should the budget be balanced every year, instead of every two or five years? The "unbalanced-budget" plan suggests that the balancing period should be roughly over the length of the cycle—running a deficit during depression years while spending to increase purchasing power and promote prosperity, and then paying off in prosperity the debt so incurred. The primary difficulties

⁴ Chapters 48 and 49.

with such an unbalanced-budget plan lie in applying it properly, and in the danger of assuming that cyclically unbalanced budgets alone necessarily assure recovery and make other remedial measures unnecessary.

In depression it is desirable to increase the monetary demand for goods and services, because such an increase will *tend* to lead toward a cumulative revival. The government therefore borrows funds (by selling bonds) and spends the funds by giving relief, carrying on WPA projects, and so on. If this process takes funds from persons, banks, or other institutions in whose possession they would have lain idle and gives the funds to persons who will spend them, it thereby increases the current volume of expenditures and the monetary demand for goods and services. But this is true *only* if those who receive the money spend it faster than would the bondholder had he not bought the security—that is, only if velocity of circulation is increased. As recovery progresses it is expected that the government would increase taxes and pay off the bonds, thus bringing the budget back into balance over the longer period.⁵ With this simple picture as background the budget problem may be examined in more detail, first as to possible depression revenue sources and then as to depression expenditure policies.

1. Depression revenue sources. In depression the government finds itself in a dilemma—its revenues from taxation fall off sharply and its expenses increase greatly owing to the necessity of taking care of people who are out of work. Unless it is willing to let people starve, it must increase its revenues somehow. One possibility is increased taxes—this source has often been used in the past. But, as unbalanced-budget advocates point out, this is precisely the time when it is undesirable to increase most taxes. Many taxes are costs to businesses, and the basic trouble is that costs are too high relative to expected revenues. Examples of such tax costs are pay-roll taxes, excise taxes (taxes on playing cards, cigarets, and other individual articles), and sales taxes, to the extent that they are borne by sellers. Such taxes are so much per unit of output or per employee, and directly increase marginal costs. They lead to a *restriction* of output and employment, not an expansion. If the consumer bears these taxes the result is equally bad—in that case the purchasing power of the consumer is reduced because of the higher prices. Moreover, such taxes are largely on goods which the poorer and middle classes consume heavily; and it is the buying power of these groups which must be revived. There are probably few worse ways to increase revenues in a depression period than by the use of such new sales, excise, and pay-roll taxes as were adopted in the 1930's. Budget-balancers advocating the imposition of such taxes in depression deny the basic claim of the unbalanced-budget proponents—that the way to obtain recovery is by borrowing and spending in order to increase monetary demand—and they

⁵ More recently one group of economists have advocated permanent deficit financing because they believe that such government investment is necessary to maintain full employment in our "stagnant economy." This argument for long-run "compensatory" spending is considered in Part X, *The Public Economy*.

disregard the deterrent effects on production. Other taxes, while likewise nonexpansionary depression financing sources, are somewhat less objectionable. Thus an income tax on high incomes might fall largely on funds otherwise "idle," so that funds so obtained *would* increase velocity and total monetary demand when spent. But it is by no means sure that the tax would fall only on "idle" funds. Many observers believe that high income tax rates and taxes on business profits are an important factor in explaining the depression failure of private investment to revive. This deterrent factor is clearly of importance, and the present tax structure especially penalizes the taking of business risks. Just how important such deterrents are in the total picture is uncertain, but that increased taxes are poor depression revenue sources is amply clear.

If we insist on raising additional depression revenue by taxation, a minimum requirement should be to avoid taxes on output and employment (on marginal costs), and utilize as far as possible taxes that have no such direct effect in restricting output. But many writers argue that a more sensible plan would be to *decrease* taxes at such a time, especially taxes on output and employment. This would mean a *decrease* in business costs, and an *increase* in purchasing power, since individuals and businesses would then have the funds to spend that they otherwise would have had to pay in taxes. Further, this approach would avoid the deterrent to new investment imposed by high taxes on profits. But in spite of these advantages of reducing taxes in depression periods, the problem of where to obtain the revenue needed to carry on relief and regular governmental functions would remain.

In view of the disadvantages of added taxation in depression and of the need for added revenue, unbalanced-budget advocates argue that additional funds should be raised by borrowing. Borrowing to finance depression expenditures will, they argue, have two major results. First, it will provide for relief expenditures and regular government functions. Second, it will "prime the pump," setting recovery under way by increasing monetary demand for goods and services. The second claim rests on two tacit assumptions. First, it is assumed that the unbalanced budget does increase total monetary demand and incomes, an assumption true only if the borrowed funds would not have been spent had they not been obtained and used by the government. Yet this premise is probably for the most part correct, especially since such a large proportion of government bonds are bought by the banks and high income groups. If the bonds are sold to banks, new credit is created; if they are sold to holders of "idle" funds an increase in velocity results. In either case, the result is an increase in the volume of expenditures and incomes. The second assumption underlying the "pump priming" proposal is that the recovery once under way will become cumulative. There is some truth in this proposition, to be examined presently, but many unbalanced-budget advocates overestimate the possibilities of their proposal. Deficit financed spending cannot alone assure recovery and prosperity.

A third possible method of financing unbalanced budgets in depression

is simply the issue of new money. Indeed, if monetary-fiscal policy were to be implemented by the issue and withdrawal of money, it would be put into circulation when needed primarily in the form of recession or depression expenditures. The possibilities and disadvantages of this method of financing have already been discussed in connection with monetary policy. Closely related and avoiding some of the difficulties of the new money method would be the financing of spending by direct credit creation by the Federal Reserve through purchase of securities directly from the Treasury, an alternative considered more fully later.

2. Depression expenditure policies. Once revenues are obtained, the problem remains how best to spend them. The two primary possibilities are (a) direct relief or dole, and (b) public works such as WPA. There are strong arguments for and against both methods.

Advocates of direct relief point first to the lower cost of that method and its direct stimulative effect, and emphasize certain drawbacks of public works. Public works projects are frequently very slow to get started after the need for relief and increased purchasing power arises; and they are very difficult to stop promptly once the need for them is past and it is desirable to cut down on public expenditures lest the continued expenditures speed recovery too fast. This is true not only because of the physical nature of the projects, but also because of the political nature of a public works program. Once such a program is started, it becomes largely a matter of political pressures and logrolling as to how many projects there shall be and where. Every locality demands its political due, and the results of the logrolling and pork-barrel methods of Congress are too well known to need description. By comparison, it is likely that direct relief would be somewhat more flexible.

Second, public works projects are likely to keep high exactly those sticky costs that ought to come down if real recovery is to be achieved. Examples are wage rates of many types of strongly unionized skilled workers and prices of monopolized building materials such as cement and steel products, costs that fall much less than most prices in recession because of special monopolistic supports. If public works programs are then instigated that lead to a large demand for such labor and such construction materials, these public works will be keeping up the very prices (costs) that should be brought into relative adjustment with other prices. Public works programs in the past have often been open to this criticism. The stated policy of paying "prevailing wage rates," whether these "prevailing" rates have fallen from prosperity levels or not and even when most workers are taking lower rates, is a case in point.

Lastly, public works may be in direct competition with private industry. For example, some persons feel that slum clearance would be an excellent type of public works, but that this would be in direct competition with private business in the building industries. Therefore, it is argued that such depression policy might restrict private investment rather than encourage it.

On the other hand, there is much to be said in favor of public works and

against direct relief. One of the most important arguments is on "non-economic" grounds. A "dole" is likely to be destructive to the morale of unemployed workers, especially if it continues for any length of time. Self-respect is lost; idleness becomes habitual; skills become rusty or are lost. Such social losses are a major cost of depression and unemployment at best; and they are likely to be much less when the unemployed are given work relief. Sociologists emphasize very strongly the destructive social effects of the "dole."

Second, many economists argue that public works are a more direct stimulus than relief payments to durable goods industries (such as construction, steel, and highways), where the depression centers, and are therefore a more powerful stimulus to recovery than simply putting relief funds into consumers' hands would be. This argument may be a strong one in favor of public works—revival does depend largely on recovery in the durable goods industries. On the other hand, such funds are much less certain to be respent promptly by the recipient than are relief payments direct to needy persons.

Another advantage claimed for public works as against direct relief is that "we get something to show for our money with public works." This is true, of course, but it is by no means decisive. The question is, do we get enough to merit the additional cost of public works over direct relief? The justification for public works in depression is twofold: first, their value in their own right as public works, and second, their value in providing relief and stimulating revival. A public works project which is not "economic" in good times may become worth undertaking in depression when the relief-revival aspect is added to the worth of the project itself. Roads, streets, public buildings, parks, schools, music and art projects—these and many others give important and lasting public gains in addition to their relief-revival aspects. On the other hand, care must be taken lest public works projects turn out to be "white elephants." Many a small community now finds itself severely burdened by the cost of maintaining expensive public buildings erected by the federal government during the depression of the 1930's. Such public works projects in their own right have a negative value to the community and must depend for their justification solely on their relief-revival advantages. Unfortunately, in the political pork-barrel methods of allocating public funds, very often little care is taken to investigate the future needs for, and costs of, projects undertaken, and "the something we have to show for our money" is sometimes primarily a continued drain on local taxpayers.

Some persons have proposed that instead of either relief or public works, the government give outright subsidies to producers in those particular industries that most require support. Such subsidies would make it worth while for producers to expand output and take on more workers, thus obtaining the same result as sought through relief and public works. The use of direct subsidies has not been looked upon with much favor in this country, and the political aspects of such a policy might be most unfortunate. The granting of direct subsidies would be an open invitation to

high-pressure lobbying for subsidies, political favoritism, and outright graft, unless some automatic basis for the granting of subsidies could be devised. Economically, on the other hand, the proposal has considerable merit in avoiding some of the disadvantages of both public works and direct relief, and it probably deserves more attention than has been given it as a possible antidepression measure.

Timing of fiscal policy

As in the case of monetary policy, the question of timing fiscal measures is of correlative importance with the measures themselves. When should government "spending" be undertaken to check recession and when shaded off in returning prosperity? If the government spends heavily in depression and recovery begins, too-long-continued spending may lead to an inflationary boom that culminates in another recession and depression. On the other hand, if government spending is chopped off too suddenly as recovery occurs, the fall in monetary demand may check the recovery and throw the economy back into depression. This is partially what happened in 1937. After several years of large government deficits, recovery seemed well under way with production again at high levels and federal government very sharply reduced its expenditures. At about the time there was much agitation for a balanced budget. In early 1937 the new Social Security pay-roll taxes went into effect, exerting a further deflationary pressure by raising entrepreneurs' costs and decreasing wages paid out to employees. In addition labor strife was great and wage rates were pushed up rapidly in certain basic industries. This combination of decreased monetary demand and rising costs led to a very sharp recession—the drop was even sharper than that following 1929, although it did not go as far. This instance illustrates first the importance of proper timing for government spending, and second the manner in which a variety of factors at work must be considered to explain business conditions at any time. Fiscal policy is only one of many important factors.

It is agreed among most economists that some government inflationary pressure is desirable in times of depression. But this does not mean that deficit financing is a cure-all. The decade of the 1930's indicates clearly its weaknesses. Deficit financing is generally a factor tending toward recovery, but it may well be offset by other factors. For example, following 1933 the New Deal administration succeeded in stimulating a rise in the price level. This, coupled with returning confidence and some sticky costs, was an important factor leading toward recovery with increased output and employment. But later wage rates began to rise very rapidly in many industries, especially under the pressure of the newly organized C.I.O. and under encouragement from the administration. To the extent that wage rates rose as fast as selling prices, the advantage of higher prices in leading to increased employment in these industries was largely lost. At the same time there developed many relative cost-price maladjustments—the cases of high construction wages and construction

materials costs were prominent. Such maladjustments check the tendency toward recovery and full employment in the industries affected; and if critical industries such as steel, railroads, and construction are involved, the resulting restraints may check the whole recovery movement. Expansionary pressure through deficit financing is only one element in a complicated set of determinants; it cannot bring about recovery unless other conditions are favorable.

Fiscal policy and the government debt

Deficit financing to combat depression is likely to lead to a large public debt if the depression continues for several years. Many persons have become alarmed during such periods at the increasing size of the government debt, fearing "national bankruptcy" or lesser catastrophes. To the extent that people are so alarmed, whether the alarm is "rational" or not, this acts as a deterrent to further recovery. Although government spending provides an impetus toward recovery, if the deficit to finance the spending creates alarm and uncertainty in the minds of investors and businessmen this deterrent may partially offset any expansionary result of the spending.

The fear that increasing the public debt will make the nation go bankrupt is almost completely fallacious. The public debt is owed by the United States to the United States, with the minor exception of a few securities held abroad. Paying it off will be a matter of payment from some people in the United States to others in the United States. Taking the nation as a whole, payment of principal or interest is merely a matter of transfer of income or wealth from some persons to others. There may even be little real transfer in paying the interest or principal. If the government should obtain revenue to pay the interest by taxing the same groups who own the bonds, the transfer is purely nominal, though actually the paying of interest and eventual payment of principal probably involve a considerable real transfer of wealth and income. It is, of course, quite conceivable that a government debt may become so large as to lower the credit standing of the government among investors—at the extreme that the government itself (as contrasted to the nation) might go bankrupt, as did the German government after the war. But the possibility of bankruptcy of the federal government here is so remote as to merit little discussion. United States government bonds are still considered the ultimate "gilt-edged" investment and even today the government obtains funds at very low interest rates.

The real objections to continued increase of the public debts in depression are the detrimental effect on business confidence and the transfer effects of paying interest and principal when due. By 1942 the interest on the federal debt approached two billion dollars annually; since then it has risen rapidly. To raise this revenue requires heavy taxes somewhere, and although the tax revenue is paid out again to securityholders it may still be a serious burden on the taxpayer, especially if the taxes fall heavily upon the poor or upon marginal costs of businesses. If the

taxes to raise interest funds fall on business costs or reduce monetary demand, a large public debt has a direct effect in checking output and new investment, as well as an indirect effect through destroying business "confidence." Whether or not further increases of public debt are desirable at any time should be determined on a weighing of all these various conflicting considerations. In any case, experience has shown that absence of the pressure of making expenditures conform to regular government receipts usually leads to growing waste in government administration. This is a homely but important argument against too easy reliance on continued deficit financing.*

Policy Toward Costs

The importance of sticky costs and particular relative cost-price maladjustments has been emphasized in connection with cyclical fluctuations. Sticky costs are essentially a monopolistic phenomenon—in a purely competitive market prices of goods and services would move freely in response to demand changes. Four of the most important types of sticky costs are (1) monopolized wage rates, (2) monopolized raw materials and semifinished industrial products, (3) public utility and railroad rates, and (4) many taxes. All constitute important costs to many types of business enterprises. The first two are sticky primarily in a downward direction—they stay up in recession when other prices fall, but they rise with other prices in recovery periods. Public utility and railroad rates and taxes have usually lagged substantially in both directions. When monopolized or partially monopolized prices, whether of labor or of products, constitute costs to other industries, their effect in restricting output is clear; when the monopolized products are consumers' goods, the restriction is directly on consumer purchases. Monopolistic restraints result in a restriction of output and employment both in the monopolies and in other industries. And this restrictive effect is likely to be particularly serious in periods of recession when selling prices are falling rapidly but costs are held high by monopolistic pressures. A similar effect results in the case of utility and railroad rates and taxes, since they too constitute important costs to most producers. Whether the market conditions approximate simple monopoly, duopoly, oligopoly, or another monopolistic pattern, or whether government policy determines rate changes, resulting sticky prices and costs may have similar cyclical effects.

What ought to be done about sticky cost-prices? The most sweeping step would be to eliminate the great fluctuations in monetary demand that makes the sticky costs of so much importance. Sticky costs are essentially contributing factors to cyclical fluctuations, not inaugurating or motivating causes. But this does not mean that no grounds exist for an attack upon the sticky costs themselves. The stickiness of many taxes and utility and railroad rates could be lessened directly by govern-

* A more detailed discussion of the effects of a large public debt will be found in Chapter 50.

ment action since all are directly controlled by government or government agencies, and there seems to be little reason why these prices should not be made more flexible. This would mitigate the seriousness of price declines for businesses concerned. Monopolized industrial prices and wage rates call for direct attack, on grounds of facilitating the allocation of resources in accordance with consumers' demands as well as in order to lessen the severity of cyclical fluctuations. Monopolized high prices in particular industries such as aluminum, cement, steel, lumber, and heavy electrical machinery,⁷ may have a far-reaching deterrent effect on production and employment, in addition to the reduced output and employment in the monopolistic industries themselves. Such prices are costs to many industries, directly or indirectly, and high costs restrict output and employment in all these other industries. Wage rates in certain crucial industries hold positions of corresponding importance. Monopolistically maintained union wage rates in the construction industry, for example, mean not only fewer jobs for workers shut out of the unions, but higher building costs and less building, which in turn means less demand for workers in all the industries connected with construction projects.

While the case for attacking particular sticky monopoly prices to eliminate restrictive price-cost maladjustments seems clear, the question of changes in all wages (the wage level) as a device for stimulating revival or checking recession is much more complex and far less decisive. Proposals usually go to one of two extremes. One proposal is to reduce all wages sharply, the other to maintain all wages high or raise them; both schemes are advocated to prevent recession and stimulate recovery.

The first group argues that a lowering of wage rates at the outset of recession would cut costs and stimulate business activity, new investment, and employment. But this overlooks an important fact, namely that if *all* wages were cut sharply when recession begins total wage-payments would probably be drastically reduced. This reduction in income and in consumer demand would probably lower selling prices and sales turnover, so businessmen might not be any better off than before nor would unemployment be prevented. Moreover, such an overall wage cut might well have such serious psychological effects as to lead to increased uncertainty and hoarding, and to further curtailment of investment. Thus business revenues might fall even more than in proportion to the wage cut, leading to decreased rather than increased employment. On the other hand, if such a wage cut was made in the trough of depression it seems somewhat more likely that the lowered wages might lead to more workers being taken on. Although wage rates were lower, the total wages paid out might well increase. What the effect would be would depend upon how many (if any) additional workers were hired at the

⁷ These particular industries are cited because of their basic importance to the economy and because all have recently been cited by the Antitrust Division of the Department of Justice.

lower wage. *But in any case the proposal is purely academic*, since there is no conceivable political possibility of obtaining such a uniform wage cut under existing institutional conditions. Actually such an attempt would probably lead to cutting those wages that had already fallen most, since those are the ones least protected by union pressures. Those wages that were relatively high would be best able to resist the cut.

The counterproposal is that all wages be maintained high at the beginning of recession or be raised in depression to increase purchasing power. It is true that *if monetary contraction were checked at the beginning of recession*, continued high wage rates would probably be desirable on "psychological" grounds, although even then *relative* wage adjustments might be necessary. But once monetary demand has decreased and selling prices fallen, the maintenance of high wages means unemployment. When revenues fall below costs, and curtailment of output sets in, then maintenance of high wages may well make the recession and unemployment more serious. With wage rates maintained high, workers are laid off and the total wage bill decreased, which in turn leads to a further contraction in monetary demand and to still more unemployment, and so on cumulatively. It should be clear that the monetary and fiscal policy adopted by the government plays a critical role in determining the effect of different wage policies in recession.

The argument that wages ought to be raised in depression to increase purchasing power is almost completely fallacious. At such a time, higher wages (costs) would mean that entrepreneurs would be even less likely to make new investments and expand output. At such times every effort must go toward encouraging new investment; and raising the most important cost is hardly the way to accomplish this end. Failure to see this simple central fact vitiates many of the "purchasing power" plans for recovery so common in recent years.

Miscellaneous Policies

All actions and policies of the government have some effect on business conditions insofar as they affect at all the expectations of entrepreneurs, investors, or consumers. A loan to Russia may make businessmen fear the coming of communism; the sympathetic attitude of government officials toward organized labor may make employers pessimistic about future profits. Any action that tends to further the feeling of certainty and confidence in the future will promote greater investment and better business conditions; if uncertainty and pessimism are increased, business plans are revised downward.

Numerous other particular policies have been suggested in attempting to mitigate business fluctuations. Of major importance are certain policies connected with tariffs and exchange control in the field of international trade and finance, but discussion of these must be postponed. The NRA constituted a combination of (largely self-contradictory) proposals for recovery. The Townsend Plan and the California "Ham and Eggs" Plan to achieve prosperity through large pensions to old

people have received widespread popularity, in spite of their very superficial economic analysis of the problems to be faced. Taxes on "idle" money and bank deposits have been proposed as devices for increasing velocity and stimulating business and employment. Everyone has a plan for recovery. But the policies mentioned in the preceding sections constitute the basic control methods now at hand.

"Long-Run" Stabilization

Business fluctuations center in fluctuating investment activity, which in turn can be traced largely to fluctuating profit expectations. Booms and depressions represent largely cumulative waves of expansion and contraction in monetary demand ($MV + M'V'$), coupled with many sticky costs, though it is clear that business fluctuations are closely related to the jerky rate of technological progress. Obviously the fundamental remedy would be to prevent these great fluctuations in investment by removing the conditions that lead to them. Though many of the causes lie deeper than the monetary changes, elimination of the monetary fluctuations is urged by many as the greatest practicable single step toward long-run stabilization.

Such economists generally have argued that the government should undertake a policy of long-run monetary stabilization at a time of relatively full employment. If, beginning in such a period, the government (or some monetary and fiscal authority acting for it) were to take positive action to prevent any substantial fluctuations in total monetary demand, it is clear that such violent booms and depressions as have occurred in the past could not recur. Such monetary-fiscal stabilization would not prevent all business and employment fluctuations, but it would prevent their attaining the major magnitudes of the past, for cumulative monetary expansion and contraction have always played a central part in the upward and downward spirals.

Such a long-run monetary stabilization policy would be aimed to *prevent* serious business fluctuations, not wait to *correct* them once they have occurred. It is argued that booms and depressions are easier to prevent than to "cure," and that the only hope for substantial mitigation lies in maintaining stability rather than in such short-run opportunistic measures as have characterized past attempts to "cure" depressions. Numerous particular "rules" or "guides" for long-run monetary stabilization have been proposed, but all aim at the same fundamental purpose—elimination of the cumulative waves of monetary contraction and expansion that play a central part in booms and depressions. Perhaps simplest is the proposal that a certain price level (let us say an index of wholesale prices) be maintained constant. Changes in the price level would give a rough indication of changing business conditions. If recession were to set in, prices would soon begin to fall, which would be the signal for government action to undertake expansionary action; if prices rose above the prescribed level, the inflation would be a signal for application of deflationary pressure. Proponents of such a stabilization plan argue

that this would give an "automatic" indication of when government action to check an impending boom or recession is needed. They suggest that it not only would remove the need for reliance on discretionary decisions by government officials as to when monetary expansion or contraction should be undertaken, but it would also act as a real safeguard against price inflation. In addition, by providing a stable value of money, such a policy would provide a much more certain basis for entrepreneurs' plans. Such certainty is important in promoting business stability.

The other definite long-run policy "rule" most often mentioned is stabilization of money demand itself (that is, stabilization of $MV + M'V'$). Some rough index of changes in total monetary demand would have to be obtained (perhaps total check clearings for the nation) and monetary and fiscal policy directed toward stabilizing this index. There are important differences in the plans, but the essential point of both is the elimination of cumulative monetary fluctuations.

One of the most serious problems faced by such a long-run stabilization policy would be the question of implementation. If the government decided to hold the price level stable, how could it do so? First, closer control than is now available over credit creation by the banks would have to be obtained, since it would be necessary to regulate closely the total amount of money *and* credit. Since there is no method of close control over V and V' , changes in velocity would have to be offset as far as possible by changes in M and M' . Specifically, to check undesired upward movements, the government would have as a major weapon the possibility of imposing heavier taxes to reduce incomes and expenditures, thereby reducing total monetary demand and checking the inflationary boom. To check tendencies toward recession, it would have the power to issue and spend new funds or to borrow and spend as during the 1930's. The problem of maintaining monetary stability would not be an easy one, but its proponents argue that it would be much easier than to "cure" depressions once they have occurred.

Elimination of monopolistic sticky cost-prices is likewise urged by many writers as a long-run policy, both to mitigate business fluctuations and to facilitate better resource allocation. It is sometimes argued that if *all* prices and costs were completely flexible, changes in monetary demand would have few serious deterrent effects. However, the more critical need is the removal of *particular* price-cost maladjustments in particular areas where they exercise a broad restraint on employment and production, and without such action monetary steps to increase employment may be nullified. Monetary and antimonopoly actions in this regard are clearly complementary. Sticky costs are cyclically dangerous primarily because total monetary demand does fluctuate violently, and elimination of such monetary fluctuations would substantially lessen the danger from such cost-price maladjustments, but this in no way detracts from the advantages of removing such individual price maladjustments.

"Business Cycle" Policy in Wartime

Where attempts to restore "prosperity" by various means succeeded only partially from 1930 to 1940, the enormous program of national defense expenditures beginning in 1941 not only succeeded in moving the country rapidly toward full employment but soon led to serious shortages of labor and capital goods in many lines in defense industries. Defense expenditures made depression deficits seem insignificant. Almost overnight the economic problems of the nation shifted from the necessity for increasing total spending to stimulate recovery to the necessity for curtailing it to prevent inflation as swollen consumer incomes bid for scarce goods. In its early phases, the huge defense effort meant a sharply increased scale of living as the unemployed were drawn back to work and incomes rose everywhere. There were both "guns" and more "butter." But soon the fundamental dilemma of a wartime economy appeared—we could not have both "guns and butter." Our resources, bountiful as they were, were insufficient to fulfill defense demands and continue to maintain the same level of civilian consumption and investment.

Wartime fiscal policy is discussed in detail in a later chapter, but its relation to business cycle policy may be briefly indicated, especially since war finance is simply the extreme case of finance under *full employment*. So long as there are unemployed resources capable of doing the work there is no reason why a war program cannot be carried out without reducing existing levels of civilian consumption. So long as there are usable unemployed resources, rearmament spending can have exactly the same effects as public works or direct relief pump-priming—that is, it can draw new resources into work without reducing employment in civilian industries. However, once the limits of productive capacity, available labor, and raw materials are reached without completely satisfying both civilian and defense demands, the choice must be made—guns or butter? Placing war production first then means that no matter what financing method is used civilian consumption must suffer the burden of giving up goods and services it otherwise would have obtained.

Actually, of course, shortages occur in many lines long before full employment of resources is reached. By the summer of 1941, acute shortages of aluminum, steel, and other such essential materials for defense production had become evident. Machine tools had months before run short, as had many types of highly skilled labor. Thus, while there were still perhaps 5,000,000 people looking for jobs, it became necessary to impose sharp restrictions on the use of such resources for civilian consumption. Among the first articles to be restricted in civilian use were aluminum utensils, automobiles, and refrigerators, and the list grew rapidly, while there remained a large reserve of unemployed resources. An unskilled worker does not help much in running a complex machine tool; large reserves of cotton do not help much in easing the shortage of steel. As long as unemployed resources remain available

there is much to be said for attempting to draw them into employment to satisfy nondefense needs, if they cannot be used in producing arms. But the fewer the remaining unused employable resources, the smaller becomes the possibility of concurrently expanding civilian and defense production and the less the case for deficit financing. Wartime inflation comes rapidly with deficit financing long before "full" employment is achieved, and failure to recognize this fact may make even strong counter-inflation steps too late when once the situation is recognized.

What will happen when peace returns and we want to shift back to a "peace" economy? Are war expenditures a permanent or merely a temporary form of pump-priming? Unfortunately, war expenditures are likely to prove only a passing form of relief from unemployment, though this is not necessarily so. With war expenditures we achieve full employment by drawing resources into war industries such as armaments, and into the industries producing raw materials for war industries proper. But with the return of peace the demand for war materials presumably largely vanishes, and the demand for workers in these industries vanishes with it. The problems of readjustment from a war to a peace economy have always been great, even when wars were small and unmechanized. Today, when entire economies are shifted to production for the vast needs of modern warfare, postwar readjustment will involve a complete shift of industrial organization and employment. When war demands vanish, a tremendous reallocation of resources will be required.

With the shift back to a peace economy, if it comes, after a strong postwar buying wave we may expect either large-scale unemployment or a vast program of public works to put unemployed workers and ex-soldiers into peacetime occupations. The immediate postwar problem is likely to be inflation, as buying power comes onto the markets before resources can be reconverted to civilian production, followed by the danger of a large-scale slump once consumer "backlogs" are filled. The great advantage of regular public works over arms expenditures is that regular public works temporarily employ workers in the hope that they will be drawn off into private employment, while arms expenditures draw workers into jobs that must be held until such time as the war demand ends, when the jobs vanish. Full employment based on war expenditures carries the seeds of its own destruction unless the war expenditures are to be a permanent part of our national life.

Part X

THE PUBLIC ECONOMY

PREAMBLE

WHAT is the proper scope for government action in economic matters? The influential nineteenth-century English liberals insisted that the function of the state should be primarily to set up and enforce certain "rules of the game" under which private enterprise could then be counted on to bring a maximum of production. Inequitable distribution of the goods produced was recognized as a pressing problem but one on the whole secondary to that of maximizing production. Another point of view has been vigorously advanced by the followers of Karl Marx, who argue that "private capitalism" must and will give way to a socialist state wherein all production and distribution will be controlled directly by the state, for the benefit of the masses.

The present attitude toward the "public economy," as the sphere of direct government control of resource use is sometimes called, lies somewhere between these two positions. Most Americans apparently still believe strongly in capitalism and the virtues of a free, private enterprise economy. Yet the scope of government action beyond the regulatory duties prescribed by *laissez-faire* advocates has grown steadily and rapidly. More taxes have been collected and more services provided by governments. Public responsibility to care for the needy and unemployed and to mitigate business fluctuations has been increasingly recognized. And now war has swept aside traditional peacetime patterns of private resource use, channeling over half the national income through governmental hands into war production. •

With half the national income flowing through the public economy, it is no longer possible to treat "public finance" as a minor and more-or-less independent branch of economics. How funds are raised by the government and on what goods and services they are spent are questions parallel in importance to the same questions applied to the private sector of the economy. This is the task of the present Part, to discover how the flow of resources and income is channeled through the public economy and how well this flow conforms to the various goals set up for the behavior of the economic system, public and private.

CHAPTER 46

Introduction to the Public Economy

THE activities of government and private enterprise are so closely intertwined at many points that it is impossible to draw a clear line between them. Government action of various types has thus far been considered in connection with the study of the private enterprise economy—in the control of monopolistic practices of business and labor, in wage and hour laws, in the provision of relief, in the agricultural program, in control of bank credit, in government spending to mitigate business cycles, and so on. Most of the activities discussed have been attempts to set up and enforce a framework of “rules” within which private enterprise can be expected to function tolerably smoothly in producing and distributing goods and services desired by consumers. In such regulatory activities usually the use of resources by government in enforcement is relatively small, and decisions as to policy are made with money costs playing a definitely subsidiary role. The enforcement of the Sherman Act involves some expenditure, but the decision as to whether we should or should not have antimonopoly laws rests primarily on the regulatory effects of the laws, not on the money cost of enforcement. But there is another quite different type of government activities, in which we shall be primarily interested in the forthcoming chapters—activities that directly involve large expenditures and whose undertaking turns primarily on comparison of the money cost of acting with expected direct returns from the expenditures. Whether or not the government builds a bridge in depression depends on the cost of the project, on its expected usefulness, and on its potency as an employment-stimulating project.

Every government activity, regulatory or otherwise, influences the allocation of resources to some extent. We may, however, set off the “public economy” as *that sphere within which government directly influences the use of resources through using them itself or through directly redistributing money income*. In contrast with this, in the “private economy” the allocation of resources is primarily through the action of private enterprise with only incidental government influence through regulatory activity. This division is necessarily very rough, but it may be sufficient to give some idea of the scope of this section on the public economy. To state the division in another way, government activities with respect to economic life for the most part fall into three classes: (1)

regulation of private enterprise; (2) direct production of goods and services—schools, public works, etc.; and (3) redistribution of money income.¹ In discussing the public economy we are primarily concerned with the latter two; the first is a part of the public economy only insofar as it involves direct government use of resources in making and enforcing the regulations.

Government Expenditures and Revenues

Expenditures and revenues in recent years

The importance of government activity in the total economy cannot be accurately measured. Obviously it is impossible to compute the monetary significance of enforcement of the Pure Food Act or of the Wagner Act or of the Sherman Antitrust Act. Yet certain significant comparisons may be made. If we want to indicate roughly the relative importance of the public and private economies, one way of doing so is to contrast the flow of money income through government hands with the total national money income. In 1938, before the great expansion of war production, the national income was around \$67,000,000,000. In the same year the expenditures of all governments in this country (national, state, and local) were about \$17,500,000,000, exclusive of debt retirement. This sum was about one quarter of the national income. Or we may compare government with total employment, and government wages, salaries, and interest payments with those for the entire economy. In 1929, before the onset of the depression with its increase in public employment and relief, total government employment (including public education) was 3,123,000, out of a total of about 44,000,000 gainfully employed. In the same year, wages and salaries paid out by governments were \$5,386,000,000, out of a total national figure of about \$49,000,000,000; and government interest payments were \$1,419,000,000 out of total interest and dividend payments of about \$11,000,000,000. Today, with the expansion of the depression years culminating in the unprecedented war program, employment by governments and income paid out by governments in these ways are tremendously larger than a decade ago, and in the fiscal year 1943 around half the entire national income is expected to pass through the hands of the federal government alone. The public economy, once insignificant beside the sphere of private enterprise, now bids fair to swallow the entire system under the pressure of total war, and history provides little reason to expect a postwar decline of government activity even to the relatively high levels of the 1930's. Any realistic survey of modern economic life must give to the public economy a place of importance parallel to that of the private enterprise economy. With half the economy's resources used directly by the government, taxes, borrowing and government expenditure policies require vastly more attention than ever before.

A brief factual picture of the sources of public funds and of the ways

¹ Obviously these categories are not mutually exclusive.

in which they are spent is essential in considering the public economy. To this end two condensed tables and a chart are presented, showing expenditures and revenues of all governmental units—federal, state, and local—for 1938, the last year for which complete data on state and local units are available. Although postwar conditions are unpredictable, prewar figures present a perhaps more “normal” picture than would current war-period data, even if satisfactory current data were available. Table 46—1 shows expenditures by major items; Table 46—2

TABLE 46—1 *

FEDERAL, STATE, AND LOCAL EXPENDITURES FOR FISCAL YEAR ENDING IN 1938
(000,000's omitted)

	<i>Federal</i>	<i>State</i>	<i>Local</i>	<i>Total</i>
TOTAL EXPENDITURES	\$7,691	\$4,358	\$6,150	\$18,199
1. Relief, welfare and social security	2,182	637	266	3,085
2. Net additions to social security reserves ..	489	516	1,005
3. Education	177	818	1,418	2,413
4. Highways and streets	260	900	510	1,610
5. Agriculture and natural resources †	1,000	73	3	1,076
6. Health and hospitals	36	270	265	571
7. Police and other domestic protection	44	138	566	748
8. National defense	1,610	12	1,622
9. Interest on debt	926	121	592	1,639
10. Debt retirement	65	135	529	729
11. All other	902	738	2,001	3,641

* Data from *Bulletin of the Treasury Department*, August, 1939. State and local expenditures in some cases are estimates.

† Includes such items as flood control, land reclamation, and so on.

TABLE 46—2 *

FEDERAL, STATE, AND LOCAL TAX REVENUES FOR FISCAL YEAR ENDING IN 1938
(000,000's omitted)

	<i>Federal</i>	<i>State</i>	<i>Local</i>	<i>Total</i>
TOTAL TAX REVENUE	\$6,034	\$3,857	\$4,920	\$14,811
1. Income taxes (personal and corporate) ...	2,762	562	3,324
2. Estate, inheritance and gifts	417	145	562
3. Property	214	4,531	4,745
4. Payroll	743	707	1,450
5. Motor fuel and vehicle	293	1,163	25	1,481
6. Liquor and tobacco	1,136	298	32	1,446
7. Sales and other excises	287	717	302	1,306
8. Customs (tariff collections)	359	359
9. Other tax revenue	37	51	30	118
Tax revenues as per cent of national income **	8.9%	5.7%	7.2%	21.8%
Per capita tax revenue †	\$46.48	\$29.71	\$37.90	\$114.09
TOTAL NONTAX REVENUE (primarily borrowing) ‡	1,657	481	1,230	3,368

* Data from *Bulletin of the Treasury Department*, August, 1939. State and local revenues in some cases are estimates.

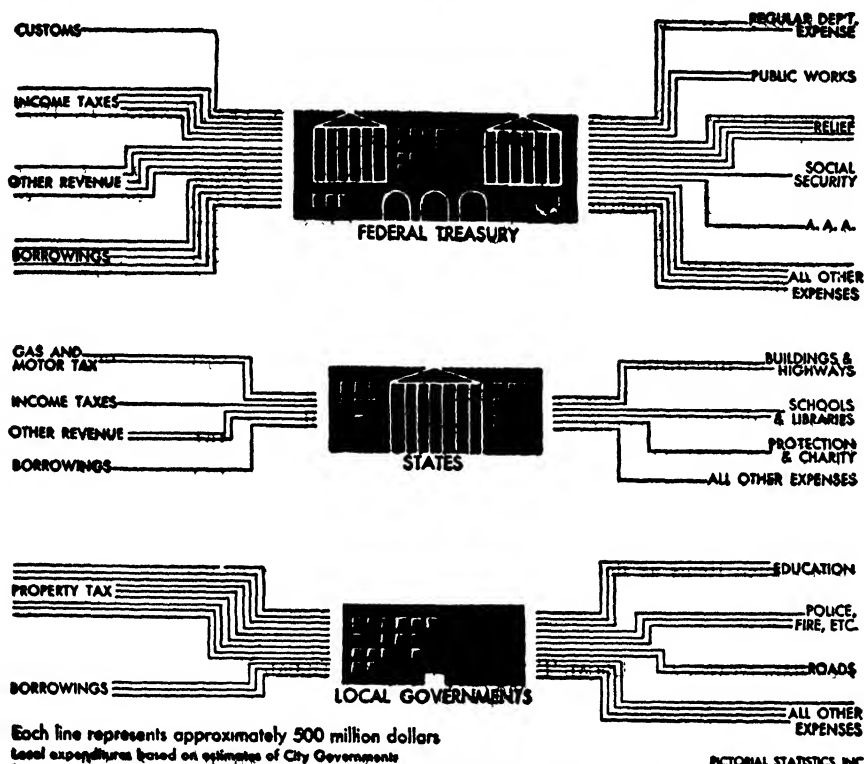
** On basis of the average (\$67,923,000,000) of the estimated national income for the calendar years 1937 and 1938.

† On basis of estimated population (129,818,000) as of Jan. 1, 1938.

‡ Excludes the income of government corporations, such as public utilities.

shows revenues from major sources; and Figure 46—1 provides a visual summary of the sources of revenues and the ways in which they are spent by federal, state, and local governments, respectively. At the bottom of Table 46—2, tax revenues are shown on a per capita basis, and as a percentage of the national income.

These expenditure and revenue figures for 1938 fail in three respects to portray accurately the present situation. First, the present huge war



Courtesy Public Affairs Committee, Inc.

Fig. 46—1. Government income and expenditures for the fiscal year ending in 1938.

program has tremendously swollen the expenditures on "national defense." In his budget message covering fiscal 1943, the President called for expenditure of \$52,000,000,000 for war alone, or about half of the estimated national income. There has been a slight reduction of other governmental expenditures at all three governmental levels, but this has been insignificant by comparison with war expenditures. Second, the federal tax burden had approximately quadrupled by the summer of 1942, with further increases in sight. And third, the fiscal year 1938 showed a much smaller unbalance between expenditures and tax receipts than the years before and since, especially for the federal government. Since our entrance into the war the budget deficit has reached unprecedented figures, with perhaps \$50,000,000,000 of borrowing expected in fiscal 1943 alone, in spite of the huge tax increases, reflecting increased

estimates of federal expenditures to \$77,000,000,000 following the President's budget message.

The expenditure and revenue systems statistically summarized in these tables will presently be considered in detail; they are brought together here for a preliminary view so that various expenditures and revenues can be viewed in proper perspective in later discussions. Unless one has some idea of the relative magnitudes of the various items, each cannot be properly viewed in its relation to other items and to total expenditures and revenues.²

The historical trend of expenditures and revenues

A century and a half ago the acknowledged functions of government were relatively few in number, sometimes hardly extending beyond the protection of persons and property from violence and certain sorts of fraud. Free schools were virtually nonexistent; much travel was over private toll roads or without roads at all; public health and government social security expenditures were rare. The years since then have seen both a great expansion in the activities of governmental units and a changing attitude toward the proper scope of governmental action. Today our governments spend huge sums on education, public health, social security, highways, relief, and so on, in addition to greatly increased sums on domestic and international protection of persons and property. But of course modern war has been far the biggest added cost item.

The increase in total federal expenditures from 1800 to 1943 is shown in Table 46—3. In 1800, the federal government spent a total of about

TABLE 46—3 *

U. S. FEDERAL EXPENDITURES FROM 1800 TO 1943 (000,000 omitted)

<i>Year</i>	<i>Expenditures</i>	<i>Year</i>	<i>Expenditures</i>
1800	\$ 11	1900	\$ 521
1815	33	1913	735
1830	15	1918	13,792
1850	40	1930	3,946
1865	1,298	1936	8,880
1880	268	1943	77,000**

* Data, except for 1943, from *Annual Report of the Secretary of the Treasury of the U. S. for 1937*, pp. 350-353. Expenditures do not include debt retirement or post office expenditures. All data are for fiscal years.

** Estimate from the President's budget message and supplementary messages to August 1942, including post office expenditures but excluding debt retirement.

² Revenues from, and expenditures on, governmental services provided on a direct sale-to-consumers-at-cost basis (such as government owned public utilities) are not included in the tables. Their omission makes little difference, however, in comparison with the total magnitudes. Although no definite figures are available, probably it is a fair estimate that total governmental sales per year on this basis are around \$500,000,000, most of which comes from the sales of a relatively few big utility projects, such as TVA.

\$11,000,000, excluding debt retirement. The present federal budget estimate for fiscal 1943, excluding debt retirement, is \$77,000,000,000. In 1800 the per capita tax burden from federal taxes was less than \$2. On the basis of estimated federal tax receipts for fiscal 1943 of perhaps \$23,000,000,000, the per capita figure will be about \$175. The per capita federal expenditure for 1943, if estimates are realized, will be over \$590. And although there are no reliable figures on state and local revenues and expenditures a century ago, the figures since 1880 indicate a steady increase there roughly parallel to the federal figures, except that in the last quarter century federal expenditures have increased much more rapidly.

These increases in expenditures and revenues in the last century and a half are phenomenal. Our governments now take from us annually astronomical sums in the form of taxes and borrowings, and return to us an amazingly large range of services and a large amount of cash payments. Why has the dollar importance of the public economy grown so? The reasons are many, but those most important may be summarized briefly.

1. First, rising prices make the present dollar collections and expenditures appear larger relative to past years than are the actual "real" increases. Since one dollar will now buy less than over most of the last century and a half, the higher dollar expenditures now do not mean correspondingly higher real income collected and spent by governments. Actually until the war price rises the value of the dollar was not recently so much less than in 1800, but the increase in dollar expenditures seriously overstates the actual increase since the period around 1880, when prices were very low. To make accurate comparisons of present and past, dollars must be deflated by a price index to convert them into dollars of the same purchasing power at both dates. Also, the dollar expenditures may overstate the actual increase in government expenditures, because in the past local revenues were often collected in the form of direct labor services (for example, the payment of local taxes by working on the roads), a practice that has now practically vanished. Thus, in the past considerable government activities did not show up in dollar revenues and expenditures, while now they virtually all do.

2. The great increases in population over the last century and a half have led to greatly increased expenditures. Even if per capita expenditures had remained constant, the increase in population from 5,300,000 to the present level would have meant about a 2,500 per cent increase in total expenditures.

3. Changing attitudes as to what constitutes a "decent minimum scale of living" and as to what services should be made available free to all (such as education) have greatly increased demands on governments. Thus there has been not only an increase in population but also an increase in expenditure per capita.

4. The increasing needs of a more complex society have continually placed greater demands on government. The growth of great cities means

an urban economy, which requires much greater government expenditures per capita than would be necessary in a rural economy. Growing demands for transportation facilities have been filled to a large extent by government action in providing highways and streets. The increasing intensity of industrial fluctuations has forced greater government expenditures on relief and public works, especially during the last few years with the increased faith in deficit financing. Such examples can be multiplied.

5. Because of its magnitude, the expenditure on wars, past, present and future, has always stood out in the federal budget, and with increasing prominence. The figures for 1865, 1918, and 1943 in Table 46—3 are striking. War has become fabulously expensive. And history shows that in each case when we have greatly increased expenditures during war periods, expenditures have failed to fall back to prewar levels after the war. War periods have proved a stimulus to greatly increased expenditures, and political pressures have made it virtually impossible to reduce to prewar levels expenditures established during war times.

Social Policy and the Public Economy

The "ends" of social policy in the public economy

In discussing the operation of the private economy, certain broad "ends" of social policy are frequently listed to provide a general directional background for consideration of public policy in economic affairs. Such "ends" are of course not recognized or weighted equally by everyone, but to the extent that they are acceptable in evaluating the functioning of the private economy they ought to be applicable to the public economy. Direct government control over the use of resources and government regulation of the private use of resources presumably are both directed to the same ends. It is useful therefore to set out again the social "ends" that were listed in discussing the private economy and to consider very briefly their usefulness as bases for discussion of the functioning of the public economy. To approach these goals contributions from both the private and public economies are required; it is the relative contributions of each that must be determined.

1. Allocation of resources in accordance with consumer preferences. Even a purely competitive private enterprise economy would fail to provide resource allocation completely in accord with consumer preferences, and such responsiveness as exists is in accordance with monetary demand; some consumers have more "votes" than others. Aside from the question of equity in income distribution and the existence of widespread monopoly, the private enterprise system falls down in a number of respects as an effective mechanism for responding to consumer preferences. Two of these flaws are especially likely to give rise to direct government production: (1) Most important is the failure of a "profit" system to produce such "public" goods and services as battleships, highways, public

sanitation, and so on; individual demands for such products cannot be effectively expressed through the price system. (2) Where social costs and social gains are not accurately reflected in money costs and money prices under private enterprise (as with the social cost of the smoke pall in a factory district), it may be that part or all of the production should be transferred to the public economy to obtain the desired allocation of resources. In all cases, the government should take steps to facilitate informed consumer choice-making, an essential part of effective consumers' sovereignty.

Yet in its endeavors to assist in the attainment of consumers' sovereignty, the government is continuously faced by the absence of consumers' monetary demand as the criterion. Unless it sells its "product," it must rely largely on our regular democratic political channels to obtain information as to what and how much of various goods and services consumers want produced through the public economy, and such information is usually vague and unsatisfactory. The problems which therefore arise will be considered expressly in a later section.

2. **Freedom in the choice of a job or business.** Attainment of this end must rest largely on government regulation of the private economy—on the establishment and maintenance of "rules of the game" that assure to individuals this democratic prerogative. The public economy itself provides only a minor opportunity for action in this direction under ordinary circumstances, unless public employment exchanges and training centers are considered primarily part of the public economy.

3. **Equity in the distribution of incomes.** Equity in the distribution of income means quite different things to different people. Insofar as it means establishing a higher "minimum" plane of living for the millions who are now very poor or redistributing income so as to reduce existing inequalities, the government can go far toward achieving this end through the public economy. Indeed, it is argued by many that only through the public economy can substantial changes in this direction be attained. The obvious method of approach is through heavy taxes on the high income groups with the funds transferred directly or indirectly (through free or subsidized services) to the low income groups, but numerous other devices are available. Any governmental activity in the public economy affects the distribution of income. The question is not whether there shall be an effect, but rather what it shall be.

4. **Progress in raising planes of living.** This aim relates to the general plane of living, as distinct from the relative positions of various groups considered in the previous point. The progress made by the government in raising planes of living depends, as in the private economy, on the availability of productive resources and the efficiency with which these resources are used. Where government production is less costly than production by private enterprise, government production is called for by this criterion. Questions that might be asked in regard to the public economy in this connection are: Does government activity encourage improvements in technology (for example, by agricultural research)?

How effectively are such improvements made available to the public? How well are resources used in the public economy? One might also ask: How well has waste, in the form of involuntarily idle resources, been eliminated? Do we have men and machines idle when people want and need jobs? The problem of maximizing planes of living is in considerable part tied up with the problem of business cycles.

5. Stability and security. Emphasis on stability and security in economic life has increased greatly in recent years. In part this involves long-period more or less "normal" needs, such as old-age insurance. But in large part it is bound up with the problem of mitigating general business fluctuations and maintaining a reasonably high level of employment and income. Government activity in mitigating business fluctuations has already been considered, but it will have to be reconsidered and amplified somewhat when viewed in the general setting of government action through the public economy.

The desire for stability and security is important in another connection as well. This is the desire to be free from the uncertainty of erratic and arbitrary shifts of government policy. Some degree of predictability of government action is clearly desired by most groups, even within a setting of widespread change and social reform. A government that frequently and unpredictably changes its rules for the private economy, its monetary and fiscal policies, its own production program, is likely to create unnecessary uncertainties and instabilities. And arbitrary enforcement of existing rules and policies is equally contrary to the general desire for greater stability and certainty.

Special characteristics of the public economy

Certain special characteristics of the public, as contrasted with the private, economy go far toward determining the "proper" and practicable sphere for direct government action in approaching any or all of the five social "ends" listed above. These special characteristics are fairly obvious, but it is important to recognize them as fundamental determinants of the scope of the public economy.

First, the state has the power of compulsion, which is necessary for regulatory action, for effective redistribution of income (by taking from some and using the funds to aid others), for restricting production and consumption of particular commodities, and so on.

Second, the state, by virtue of being made up of all its citizens, is in a position to undertake provision of many goods and services that are generally desired but not marketable under a price system in the private economy. We all want police protection, an adequate army and navy, sanitation, and so on, even though the services could not be sold to individuals on a per unit basis or the benefits even roughly measured so that taxes could be levied on a cost-of-service-provided basis.

Third, owing to its long life and great resources, the state is able to undertake long-run, expensive projects, which, though desirable, would not be feasible for private enterprise, for example, the conservation of forests

in our national parks. When the state allocates resources for preserving forests, mineral deposits, or farms (for example, passes conservation laws for oil or soil) under conditions in which it would not be profitable for private enterprise to do so, it is in effect either subsidizing or regulating present production to counteract the difference between current private costs and long-range social costs including replacement of used resources.

Fourth, the state, being free from the "profit motive" and having the power of compulsion, is able to make its revenues fit its expenditures (within limits) rather than the reverse. This characteristic is especially important in permitting attempts to promote economic stability, either by unbalancing budgets as a short-run policy to mitigate booms and depressions, or by adopting a long-run monetary policy such as price stabilization and maintaining the stable policy in spite of continued excesses of expenditures over regular receipts.

These and other considerations indicate that the scope of the public economy must be large. A tradition of trying free enterprise first and introducing government only where and when free enterprise seems to fall short, has exerted a continuous and powerful check to the expanding scope of government action. But the contributions of government have come to be regarded as more and more necessary, even aside from such emergencies as war. In the modern world, the public economy bulks increasingly large. And with the onslaught of the present war, it bids fair to swallow up the bulk of the private enterprise economic system to which we have long been accustomed.

Consumers' sovereignty and citizens' sovereignty

Allocation of scarce resources among alternative purposes is the basic economic problem. But this problem must be faced somewhat differently in the private and public economies.

As has been emphasized many times, the allocation of resources through the private economy is primarily in response to consumers' money demands for goods and services. Resources are allocated more or less automatically through the price system as individuals (both entrepreneurs and owners of productive agents) strive to maximize incomes. To the extent that consumers are "sovereign," resources will be so employed that no one of them could obtain a greater dollar reward from consumers if it were to be shifted to performing any other service. Individual consumers' wishes will be fulfilled in proportion to the number of dollars that they have to offer for desired goods and services. Actually, as is easily apparent, there are many interferences with such an adjustment.

There are some types of government activity in which the entrance of government as a producer of goods and services does not influence these statements about the allocation of resources. For example, if an enterprise is publicly owned and operated so that marginal unit cost equals marginal unit revenue, it will compete with private enterprise for resources and will obtain more resources as long as it can afford to pay more for them than private entrepreneurs can. As soon as the enterprise is unable

to pay prevailing prices for productive services, it must lose resources to higher bidders.

But cases where government services are sold to consumers on a price basis are the exceptions; in the vast majority of cases there is no direct purchase by which consumers express their preferences and direct the allocation of resources to or away from the government undertaking. Where the dollar votes of consumers are absent as guides, the allocation of resources in the public economy of a democratic country is, ideally, carried out primarily through a process of political voting. If our governments are truly democratic, we as citizens express our preferences for battleships, roads, zoos, education, and so on, through voting for our representatives; and they in turn vote funds to procure resources to produce the goods and services that we have indicated to them as important to us. Thus there ideally is a sort of "citizens' sovereignty," approached through our representative political system in the public economy, which in some ways resembles the "consumers' sovereignty" approached in response to consumers' demands in the private economy. There are, however, significant differences in the two cases. Two especially deserve attention.

First, in the private economy the voting for resource allocation is on a *one-dollar-one-vote basis*; in the public economy in a democratic country it is on a *one-person-one-vote basis*. In the private economy, the more dollars that are offered for any good, the more will be the profit in making it and the more resources will be devoted to producing the good. Since the rich man has many more dollars to spend than most of us, he has correspondingly more votes (greater power) over the allocation of resources. It is largely for this reason that yachts and caviar are produced while millions go ill-fed and clothed. On the other hand, in the public economy a democratic system attempts to provide that every citizen shall have equal vote (power) in determining government policies. Here, theoretically at least, the poor man has equal power with the rich in determining whether we should have more or fewer battleships, free schools, public works, AAA, and so on. Of course, in a nondemocratic political system this approach to citizens' sovereignty is much less close, and indeed there are serious impediments to its achievement even in our own political system.

Second, in the public economy there is no such comparative weighing of specific benefits and sacrifices by consumers as exists in the private economy. The greatest weakness of the transmission of votes in the political system as compared with that in the private economy is the absence of adequate means for weighing alternatives on the part of the various voters and for expressing their various sets of preferences specifically through the ballot. In private life Mr. X is likely to consider fairly carefully whether he prefers to spend \$10 on football tickets, an electric razor, or a new pair of shoes. In choosing one he weighs it against the other alternatives given up. But Mr. X as he votes to elect his representative, votes on a whole group of issues, and has no opportunity to

distinguish between the things of which he approves and those of which he disapproves—he votes for one complex combination against another. Even in votes on specific expenditures, for example, the expenditure of the local government on education, he has difficulty in weighing this public expenditure against unknown possible alternatives. Moreover, the voter is rarely in a position to compare a particular expenditure with its costs in the form in which taxes will be levied to finance it. Even when costs and benefits are related, they cannot easily be put on an individual basis for decisions by individual voters. John Brown votes Tom Smith's money for school lunches or a new civic swimming pool. These considerations should warn against carrying too far the analogy between "citizens' sovereignty" and "consumers' sovereignty," even though its importance as an indicator of democratic transmission of citizens' wishes with regard to public use of funds bears some emphasis. The emphasis must be on group rather than on individual wishes, and there is no common measuring rod by which alternative uses of resources may be compared and some maximum or ideal adjustment defined. The extent to which government action is consistent with the criterion of consumers' sovereignty can be quite precisely evaluated insofar as government takes part in the allocation of resources in response to consumer dollar votes. Insofar as government directs resource use in response to citizens' political votes, the measures of success are much more ambiguous and must be applied largely by the political scientist rather than by the economist.

CHAPTER 47

Public Expenditures

THE most direct approach to the functioning of the public economy is through examining the expenditures and revenues of governments. But neither expenditures nor revenues can be viewed separately if one is to obtain a balanced picture of the manner in which the public economy operates. We can say something about the economic effects of a high income tax, but unless we know whether the funds obtained are paid back to the rich through interest on bonds or to the poor as relief the analysis must be seriously incomplete. The same problem arises in every other case. However, it is necessary to begin somewhere—to study factors one at a time before they can be fitted together in the whole complex picture. Therefore, an analysis of public expenditures is undertaken first, and the means used to finance these expenditures are considered later.

The primary reason for considering expenditures first is that expenditure decisions commonly come first in the public economy. Generally our governments decide what expenditures “should” or “must” be made—then they take steps to raise the needed revenue, by taxation, borrowing, or other means. The problems of raising revenue, though of some importance in determining the scope of expenditures to be made, are often secondary. Examples are plentiful on all levels of government. Today we decide first that we must have fighting forces and war materials sufficient to defeat the Axis—then we consider ways to finance these expenditures. Local governments commonly estimate expenditures for the coming year, and then levy taxes sufficient to raise the sum needed. On the other hand, even such a basic institution as the school system suffers from reduced appropriations when hard times make funds difficult to raise.

The present chapter and the following two on taxation proceed for the most part on the assumption of relatively full employment, deferring temporarily, in the interest of clarity, consideration of the problems arising in connection with general business fluctuations and “underemployment.” The interrelations of governmental fiscal policies and the level of employment and income are then discussed in Chapter 50, “Fiscal Policy and the National Income.” Adherence to this separation is not always practicable, but it is convenient to approach governmental policies in the Public Economy in this fashion. It is therefore important to bear in mind the

assumption of relatively full employment throughout these three chapters except where otherwise specified.

This assumption of relatively full employment of resources is of special significance for consideration of government spending in the public economy, since it means that the government can use resources only by transferring them away from someone else. With full employment, if the government wants to build a new building or highway, it must draw resources away from private users who might otherwise have built new houses, tennis courts, or whatnot. Commonly this transfer of resources from private to public uses is accomplished by taxing funds away from private individuals and businesses (which reduces their spending power) and then spending the tax receipts in the market to bid away the resources needed for public use. Sometimes funds are obtained from the public by borrowing; in a few cases the government simply issues new money. But no matter what financing method is used, under the assumption of full employment increased use of resources by the government necessarily means a transfer of resources away from private users. In the stress of war, sometimes the financial overlay of this transfer is sloughed off and the government directly requisitions materials, plants and manpower; but for the most part the use of resources in the public as in the private economy is guided by money flows.

The Scope of Government Expenditures

The principle of maximizing "social welfare"

On the assumption of relatively full employment how large ought public expenditures to be? Over what fields should they extend? The answers to these questions depend on a multitude of factors, and the weight to be given each depends on the ends it is hoped that the expenditures will achieve. Fundamentally, the optimum range of public expenditure depends on the real income taxpayers and other citizens are willing to give up in order to make possible the desired public expenditures. Only through a comparison of the real income given up by taxpayers or others with the real income gained by those benefiting from the ensuing expenditure can the optimum range of public expenditures be determined. This fundamental principle follows from the fact that under full employment government expenditures necessarily involve a shift of resources out of the hands of private users into public use.

On the basis of this fact, economists have generally agreed how the range of government expenditures ought in principle to be determined: Government expenditures should be increased up to the point where the marginal loss of satisfaction to those providing the revenue is just equal to the marginal gain in satisfaction derived by those benefiting from the expenditures. As long as the marginal gain in satisfaction to those benefited is greater than the marginal loss suffered by the revenue-providers (or others affected adversely), clearly there is a net social gain in increasing government expenditures. As soon as this point of

equality of marginal gain and loss is reached, clearly there is a net social disadvantage in carrying government expenditure farther. Since "satisfaction" presumably includes all satisfactions, this is merely a statement that "social welfare," or the "total satisfactions of society," would thus be maximized. Put in other words, every resource would be used where it contributes most to "social welfare."

If society were a single unity, and government the all-comprehending brain of the unity, application of this maximizing principle would be very simple, since the government could then easily weigh satisfactions lost against those gained by undertaking various expenditure projects and could regulate its policies accordingly. This would be similar to the way in which a highly "rational" individual acts in allocating his own limited resources so as to maximize the satisfactions that can be obtained from their use. But in fact society is far from being such a unity and government is far from being an all-comprehending brain. Practical application of this principle would involve comparison of the marginal satisfactions of millions of different people affected by each contemplated policy, even if it is assumed that the government knew everyone who would be affected by the policy in question. Although some economists are fond of attempting intricate methods of applying this principle (a sample of which will be noted in the following chapter), it should be clear that any such interpersonal comparison of satisfactions given up by some against satisfactions gained by others must at best be a very rough one in practical applications. None the less, such a comparison remains the basic principle upon which the range of public expenditures—and therefore of the entire public economy—must logically depend.

The principle determining the distribution of government expenditures, as well as their total volume, is contained in the general principle of maximizing "social welfare" by a comparison of satisfactions lost and gained through public expenditures. In considering every potential government expenditure, the same question can be asked: Will the gain from use of resources in this manner be greater than the loss in drawing them away from the uses to which they would otherwise be put in the private economy? At any given time, presumably there is some order in which all possible government projects could be arranged according to the satisfaction derived from spending another dollar on each. The order in which dollars ought to be spent depends on this rank of potential satisfactions. Resources should be so distributed among all public uses that the marginal return of satisfaction from all of them is the same. This is simply a part of the general principle that, if the "total satisfactions of society" are to be maximized, resources should be distributed among all uses, public and private, so that no one unit could contribute more in any other use—but the very hazy meaning of such concepts as "social welfare" and "total satisfactions of society" for practical uses hardly needs be labored.

A further point on the level of abstract general principle bears emphasis. The extent to which public expenditure should be carried depends partially on the manner in which the burden of providing the resources

(revenues) is distributed. In everyday terms, how many dollars the government ought to spend depends partially on what sort of taxes are used to raise the revenue. If we assume that in at least a majority of cases the satisfaction lost by giving up a dollar is greater for low than for high income groups, then the more of the dollars which are spent that are obtained from the higher income groups the farther it will be proper to carry government expenditures, other things being equal. Conversely, the more heavily taxes bear on the lower income groups the sooner will the point be reached where the loss of satisfaction exceeds the gain from government spending.

Unfortunately, the government cannot simply balance at the margin each citizen's desire for highways against his desire for food in the way an individual can balance his desire for shelter against his desire for food, though it ought roughly to attempt such a comparison for the populace as a whole. But even if the government could carry out a fairly satisfactory marginal comparison of satisfactions, further difficulties would arise from the fact of compulsion in government revenue-raising. First, there are necessarily costs of administration—not merely the costs of the government in collecting and spending funds but also the costs and trouble of the public in complying with the governmental compulsion. Such costs constitute an offset against any net gain in satisfactions that might otherwise have arisen. Second, any increase in taxes may well lead taxpayers to act differently than they would otherwise have preferred to act. For example, a heavier tax on my income in Maryland might well cause me to move to another state that I should otherwise like less, thus causing me an indirect "damage" comparable to the direct damage that giving up more of my income might cause. Such indirect damage, often very important in the case of heavy new taxes, must also be considered as a net offset against any gain that might otherwise have resulted from additional government spending.

Money and real transfers in the public economy

Some government expenditures involve simply transfers of money income—others involve the actual use of productive resources in the public economy. An example of the first type is the use of tax funds to pay old-age benefits or interest on government bonds. In such cases the government only transfers funds from taxpayers to old people or bondholders. Although this money transfer may change the distribution of private command over resources in the private economy, it does not involve the use of any resources in the public economy. In the second type, by contrast, if the government uses the tax funds to build a highway or hire clerks or operate schools, resources are directly drawn from private into public use. Of course the services provided by the highways, clerks, or schools are used directly or indirectly by citizens, but the control over the resources is in the government.

Some writers emphasize strongly the distinction between these cases, arguing that the economy can stand a much larger volume of money

transfer expenditures than of real transfers of resources into the public economy. It is very easy, however, to overemphasize this distinction. If the same taxes are used in both cases to raise the funds, there is no difference in the persons giving up monetary control over resources. And unless it is assumed that funds spent directly by the government are completely wasted, the resources used in the government's expenditure (let us say, steel, concrete, and labor for building a highway) are actually distributed to the people who use the highway and to others who may benefit indirectly from it. On the other hand, if the government merely uses the funds to pay out as interest to bondholders, the bondholders have funds to obtain control over resources. Then the steel, concrete, and men may be used instead in building a factory, or the bondholders may use the funds to buy automobiles. The difference between this type of spending and direct government spending lies primarily in the different distribution of resources that follows. Only if the bondholders hold the funds idle or return them to the government in taxes or investments is the result fundamentally different, since in those cases the bondholders do not exercise their purchasing power and there is no one to bid for resources with the funds given up by taxpayers. In all other cases, the difference lies only in how the resources are redistributed—if the government spends directly it controls the redistribution of the resources directly; if it simply pays out money income without hiring productive services, the recipient of the income controls the redistribution of resources through his expenditures.

The real significance of the distinction between the two cases appears to lie more in the efficiency with which government can direct the redistribution of resources in accordance with such social criteria as may be set up. When the government directly spends the funds, it can be sure that resources are allocated to those projects that seem to it most important—to highways, to free schools, to subsidized low-cost housing, to beautiful parks, to police protection. If redistribution of income in favor of the lower income groups is an important desideratum, this can be accomplished directly by providing free or subsidized services for the poor. By contrast, if the government merely makes money transfers from taxpayers to members of the low income groups, there is considerable doubt as to how the distribution of resources will actually be affected. One might suppose that funds so transferred would then be spent on what the poor "need" worst, be it food, schooling, clothing or whatnot, but as is well known actually even very poor people may prefer to spend part on cigarets, cheap jewelry, or entertainment. Which course one prefers in shifting resources toward the lower income groups depends largely on his social views—some people feel that it is important to let poor people have freedom to spend as they please; others feel that it is the responsibility of society (of the government) to see that such persons are given the things that they most "need" if they are to be directly helped by the government. In any case, it is clear that the government must directly provide many of the services "needed" by both poor and rich—an army, navy, roads,

police, and the like, since these services cannot satisfactorily be provided through the price system.

As indicated above, one important characteristic of much of the government spending that constitutes merely a transfer of money income is that it is relatively fixed by contractual agreements, especially interest and principal on public debt. This portion of government expenditures largely goes to the upper income groups, and hence provides no possibility of effective redistribution of income to help shift resources to the lower income groups. To the extent that debt payments are financed by highly progressive taxation, it may be argued with considerable justification that the transfer merely shifts funds from one pocket to another of the same group of people and has little real effect on the distribution of resources or the level of economic activity. To the extent that payments are financed by taxes that fall on lower income groups, or on business enterprise or investment, income inequalities may be exaggerated and a repressive effect be exercised on business activity.

The government as transfer agent and as entrepreneur

To the extent that resources are directly used in the public economy, alternative courses are open to the government. Either it may simply buy products and services from producers and transfer these finished "products" to the recipients of government services, or it may act as entrepreneur and organize production itself and then distribute the finished products or services.

For example, in obtaining a new highway, the government may simply let a contract to a private builder, or it may hire workers, buy materials, rent equipment and build the highway itself.

If the principal aim is efficient use of productive resources, the government ought to act as entrepreneur whenever it can thereby get the desired result more cheaply than would be possible by letting out the contract to private producers. For any given project, the money costs of carrying out the project provide a generally satisfactory guide as to which is the more efficient way. Although more fundamentally the problem might be viewed in "real" terms, in most cases money costs are the best available working basis for a decision on the grounds of efficiency in the use of resources. As indicated when "public utilities" were discussed, industries operating in the range of decreasing costs are commonly those in which government operation or control is approved by economists. However, this reasoning stems largely from the need for preventing monopolistic exploitation of such seller positions—the question of efficiency of operation under public versus private enterprise is a separate matter and one that cannot be decided simply on principle. It is a question of fact to be determined in each individual case.

In fact numerous other considerations enter prominently into the picture in determining whether governments should act as entrepreneurs in the public economy. Most important are probably the general political-social-economic mores in favor of maintaining a relatively large area for

private enterprise, at least in peacetime. In the past the majority of people in the United States clearly preferred to have the government "stay out of business" except in a few special cases such as the public utilities. A widespread distrust of government ownership and operation of enterprises as inefficient and muddling has been evident, though the evidence on which to base such decisions is very limited in view of the special circumstances under which government operations have usually been carried on. If people feel that private enterprise is a desirable end in itself as compared with government enterprise, then the test of how far government operation of enterprise ought to go can no longer be based solely on considerations of economic efficiency in the usual sense of the term.

Whether the government purchases the goods and services it distributes or acts as entrepreneur in producing them need make no difference in the policies it follows in distributing the finished goods and services. Suppose the service is the post office system, which is a case of government entrepreneurship. The postal system may be operated to make a profit, to break even, or to suffer any loss up to 100 per cent (that is, free service to all). If the government paid private enterprises to carry on the work of the present postal system, it would still be necessary to choose among exactly the same alternatives. The government could instruct the private enterprises to charge such rates as to make a profit and then seize the profits by means of special taxes or charges; it could establish rates that would just cover the costs paid to the private firms; or it could subsidize the private firms and instruct them to charge rates lower than the cost of the services rendered. The same general circumstances obtain in other cases in which the government is not at present acting as entrepreneur.

The Distribution of Government Expenditures

The tables and chart in the preceding chapter indicate in detail the distribution of government expenditures by types of expenditures and by levels of government. If, in accordance with the assumption of relatively full employment, relief payments and a large portion of agricultural payments are eliminated, the remaining expenditures fall largely into two groups—for war and for more or less "regular" governmental services. Government expenditures may be discussed under these loose headings, deferring "depression" spending to the later chapter, "Fiscal Policy and the National Income."

"Regular" expenditures

In "regular" expenditures, the major items are education, highways and streets, police and other domestic protection (including the courts), public health and hospitals, social security, and the preservation of natural resources. These are the government-provided services which most people pretty much take for granted without ever stopping to think of their basic importance in everyday life. How widespread and critical such services are could be fully realized only if we were forced temporarily to do

without them. Suppose, for example, all roads, bridges, and waterways built and maintained with public funds were suddenly closed; that all public buildings were boarded up; that all efforts at law enforcement were completely suspended, throwing every person on his own for protection of self and property; that fire and sanitary protection were eliminated; that all public schools were closed; that all publicly supported hospitals were closed and all public health measures suspended; that all legislative bodies were abolished; that our foreign embassies and consulates were closed and all diplomatic relations broken off; that all provisions for the welfare and security of workers and of the aged and infirm were abolished. So great would be the disruption of normal life, so complete would be the confusion, that consideration of its magnitude may indicate how vital these routine regular services are to modern living.

A small-scale sample of such curtailment of government services occurred in the depth of the depression in 1933 and 1934. Lack of revenue forced state and especially local governments into drastic retrenchment programs. Schools were closed in many localities; police and fire protection were sharply curtailed; city buildings were allowed to go unrepaired and sometimes closed; streets were not cleaned and were permitted to fall into worse and worse disrepair. One city went so far as to empty its jail. In almost all cases city employees either received severe wage reductions or simply worked without pay, hoping to collect their back pay in the future. Thousands of employees were laid off. All these contractions occurred in spite of the fact that these governments continued to collect and spend huge sums. Complete cessation of government services would have been infinitely more disruptive.

Table 46—1 of the preceding chapter indicates clearly that, except for social security and the preservation of natural resources, state and local governments provide most of these regular services. This arrangement is the outgrowth of earlier days when virtually all governmental expenditures save those on national defense were carried on by the state and local governments. More recently, although the scope of federal activities has widened, many people have argued that the traditional division of functions is still the most logical arrangement because of the nature of the services rendered. It seems clear, for example, that a city is the logical government unit to finance a garbage disposal system. On the other hand, many persons argue that some of these local expenditures—education, for example—could be better financed and administered by state or federal governments. Even though a local government may be best fitted to administer (spend) funds for particular local services, it may be better to have the funds collected by superior governments if such collection is more efficient or if the burden of financing the services can thereby be more equitably allocated. The arguments favoring more centralized control over functions generally considered local or state-wide in character often are closely related to a desire to set and maintain certain minimum standards in the various expenditure fields—education and public health, for example. Actually, the present system of grants

from federal to state and local governments, and from state to local governments, goes farther toward this end than is often recognized. This important question of federal-state-local governmental financial interrelations is considered at greater length following consideration of various methods of obtaining the public revenues to be spent. On balance there seems to be a strong case for greater centralization in both revenue collections and expenditures if the aim is primarily to attain a governmental fiscal program that is efficient in maintaining minimum standards, equitable, and conducive to a high and relatively stable level of employment and national income.

Although there has been a steady growth in expenditures by state and local governments on almost all regular functions, the relative percentages spent on the various functions have been surprisingly stable over the last quarter century. In both state and local governments, expenditures on "general government" and on protection to person and property have declined somewhat in relative importance. Expenditures on education have grown substantially in local governments relative to other spending, while in state governments their relative position has remained essentially unchanged. In highways and roads the situation is essentially reversed. Local expenditures show a considerable relative decrease, but this is more than offset by the substantial relative increase in state expenditures. Federal grants bulk large in increasing total expenditures on both education and roads, especially in recent years. Expenditures on charity, hospitals, and corrective activities have increased in relative importance on both levels, but especially in state governments where they show far the largest relative increase of any item. To a considerable extent, however, this has been due to increased relief expenditures during the recent depression and is therefore more properly considered in "depression" rather than in "regular" expenditures. Aside from these major items there has been comparatively little relative change in the importance of other expenditures. The federal government, aside from its small direct expenditures and substantial state-local grants to support these functions, enters the "regular services" picture primarily through its co-responsibility with the states for the greatly expanded social security program.

War expenditures

Except for the upsurge of the 1930's, every great wave of increased public expenditures has come because of war. And following these vast increases expenditures have never fallen back to prewar levels. Today, the United States, like most of the other major nations of the world, is involved in the most devastating war the world has ever known. Over half the resources of several major belligerents are already being devoted to war—what the rest of the war years will bring is unforeseeable. Money costs have already risen to astronomical figures.

This war, like no other before it, is a war of entire economies. Until the present century wars were fought largely by small armies, often of professional soldiers. Civilian life, except in the actual military zones,

was seldom drastically disrupted. Comparatively little of business and industry had to be converted to the production of war materials, and where this was necessary it was largely on a "business-as-usual" basis, at the option of the producer. World War I, with its devastation of vast areas of Europe and its millions of armed men, provided a preview of the penetration of total war into the normal peacetime pattern of economic life. In Europe large portions of the belligerents' productive capacities were devoted to war production; even in this country a large-scale diversion to war output had been accomplished by 1918, especially in heavy industry. Agriculture provided huge amounts of foodstuffs, but this involved little fundamental change for the farmer.

That the present war is incredibly more "total" and disruptive than the last has already been amply demonstrated. Total war is tremendously expensive, in terms of both resources and money. The process of making a big gun or military plane goes back many stages through the economic system. There must be skilled workers and highly specialized machinery to make the instruments of war, and many other equally skilled workmen and specialized machinery to make the machines to make the war goods. Vast amounts of materials—steel, rubber, chemicals, silk, copper, and the like—are consumed in carrying on a single hour of modern war. All economic life must be organized to fight the war; even that part left to civilian production must be regimented because of the scarcities forced upon it.

Naturally the finished products—the implements of modern war—are very expensive in money terms. A few concrete examples are likely to be more meaningful than total expenditure figures. Seventy-five mm. field guns cost upward from \$7,500, while big railway guns cost upward from \$200,000. The latter have to be returned to the factory for rebaring after firing at most one or two hundred shells. Shells for artillery cost anywhere from a few dollars up to \$5,000 or \$6,000 for one of the largest shells or bombs. A battery of 40 rapid firing antiaircraft guns of various sizes may shoot nearly a million dollars worth of ammunition in twenty minutes if operated at top speed. Small pursuit planes cost from \$30,000 up; heavy bombers from \$100,000. Beyond the actual fighting implements, soldiers must be clothed and fed, both in training and in battle. War machines must be maintained. The sums expended on supply for an army of several million men are so vast as to be almost meaningless to the average person.

Following the last war we were flooded with estimates of the cost of the war. These ranged all the way from \$186,000,000,000 to \$492,000,000,000 for all nations involved. These estimates involved valuing each human life lost (often at around \$7,000), each battleship sunk or damaged, and so on. We were told that war expenditures by the United States were \$22,000,000 a day for the two years during which we were involved.

The essential meaninglessness of such figures should be obvious. Under such circumstances, costs can only be measured in real terms. When half

our productive resources have to be taken away from civilian output and devoted to war, how are we to measure in dollars the hunger of those who do not get enough food, the want of those without adequate housing, the dissatisfaction of those who must go without silk stockings, the trouble of those who must walk to work when they have no automobile tires, the pain of families disrupted by sons and fathers sent away to fight? How can a human life lost, or a case of shell shock, or a historic landmark destroyed, be valued? What do dollars tell of the cost of giving up all new radios and automobiles and refrigerators that we may have guns and tanks? The costs of war are real—the resources given up from civilian use in order to produce war goods, the resources forever destroyed in the devastation of war. Never does the fundamental problem of allocating scarce resources to the greatest needs stand out so clearly as in wartime. Never do money and money costs shrink into such insignificance beside the real factors. Never do the most important uses of resources stand forth so clearly so that no hesitation need exist about direct action to speed resources from nonessential into essential uses. In war the fundamental decision must be how many resources are to be diverted to war production; given this decision, the financing method used is a secondary, though important, consideration.

Obviously the costs of war do not end as the last cannon is fired. The destruction of men and of capital and consumers' goods accumulated over many years and the impoverishment of the war generation necessarily leave a mark that remains long after the peace has been signed. The present generation, as will be explained later, cannot escape the costs of war, whatever the financing method used, but modern war also leaves its scars on future generations long after it has passed. Undernourishment, disrupted families and social relationships, rundown and obsolescent civilian goods equipment, depleted supplies of vital materials, millions of youths trained only to fight—these come even in countries not swept by battle. Where actual combat has raged, physical destruction is terrific, beyond all these less violent war results. Everywhere are the problems of shifting men and industry back to a peacetime basis without disruption and depression. The present generation cannot escape the real costs of the war; neither can succeeding generations escape the lower planes of living imposed upon them because of the wars of their predecessors.

In addition to these real costs for future generations arising from all-out war, there are likely to come continued higher money expenditures for government, even though complete peace is attained. Never have government expenditures returned to prewar levels after our major wars. One reason for this failure of expenditures to fall back completely following wars arises from the fact that inflated war price levels have seldom fallen back to prewar levels. Thus for the same goods and services after the war, government expenditures are larger in money terms. Beyond this, however, war provides two major continuing claims on government expenditure. One is pension and bonus payments, especially for war

casualties and their dependents but also for uninjured members of the armed services. The other is interest payments on funds borrowed to conduct the war, and possibly repayments of principal. Perhaps most important of all, having become accustomed in wartime to spending astronomical sums with little restraint, government officials are likely to find it very difficult to readjust their thinking and political actions to prewar standards. Such legislative and administrative habituation to an enlarged scale of expenditure may gradually wear off but history attests its tenacity.

Whatever may be the ethical bases for pensions and bonuses, the facts are that such expenditures bulk large for decades following every war. The federal government still pays pensions to hundreds of Civil War veterans and their wives. The pension lists for the Spanish-American War are much longer, and those for World War I involve large sums annually. Not only direct cash pensions to war casualties and their dependents are involved; the government also maintains many institutions for the care of war veterans. In addition, there have been recurring demands for bonuses to all war veterans, regardless of lack of injuries. The pressure that can be brought to bear on Congress by such a large and influential group is exemplified by the successful demands for a bonus of over \$2,000,000,000 paid to all World War I veterans under the Bonus Act of 1935. Moreover, expenditures of this sort promise to increase greatly following the present war, in view of the now accepted government responsibility for those unemployed. The problem of returning war workers and men in the armed services to peacetime jobs after the peace will be far the greatest we have ever faced. Following a possible temporary burst of business activity as war restrictions are lifted, widespread unemployment seems likely unless positive and vigorous steps are taken by the government to support those unemployed and expedite their training and their placement in civilian jobs. Government expenditures for this purpose may be called veterans' bonuses, relief, public works, training subsidies, or whatnot. Whatever form they take, they will partake of the nature of both "war" and "depression" expenditures, though if wisely enough undertaken they may prevent the worst features of a possible postwar depression.

Looking at expenditures before and during the war, there can be little doubt that our vast "defense program" beginning in 1940 and the even vaster "war program" after December 7, 1941, proved the greatest anti-depression spending program in history. To a considerable extent, defense and war expenditures directly replaced existing depression expenditures, though some expenditures originally undertaken as "depression" measures are likely to be continued permanently on a "regular" basis—for example, free school lunches for poor children and reclamation projects. Unfortunately, however, the drastic demands of war on manpower are likely to prove only a temporary cure for unemployment because "war industries" can provide employment only so long as the demand for war production continues. Large-scale war expenditures may well

create maladjustments to be faced on return to a peacetime economy more serious than those responsible for the prewar unemployment, though not necessarily so.

The distribution of expenditures by income groups

It would be exceedingly valuable if accurate estimates of the distribution of governmental revenue collections and payments by income groups were available. Certain estimates of governmental tax burdens by income groups are available—these are given in the following chapter on taxation. But only very sketchy estimates of governmental expenditures so arranged exist, and these cover only the spending of the federal government. Unfortunately, moreover, even these tentative estimates of expenditures are not directly comparable to the more complete ones on federal taxation.

Roughly, the picture presented by the available estimates of federal expenditures by income groups is this.¹ Over the period from 1930 through 1939, the lowest third of the population (roughly, those families receiving an average of under \$780 annually) received slightly over 25 per cent of the funds paid out by the federal government; the middle third (roughly, those families receiving an average of between \$780 and \$1,450 annually) received slightly less than 15 per cent; and the upper third (roughly, those families receiving an average of over \$1,450 annually) received the remaining 60 per cent. These estimates involve allocating every type of government expenditure among the different income groups. Some expenditures are relatively easy to allocate, such as depression relief payment; others are far more difficult, such as war expenditures. In view of necessarily arbitrary allocation of most expenditures among different income groups, these figures should be taken only as the roughest sort of approximation. Yet two facts seem fairly clear. One is the importance of relief, public works, agricultural aids, and other such depression expenditures in building up the share going to the lowest third. The other is the great importance of payments on the national debt, on regular departmental functions, and on the army and navy in accounting for the major share going to the highest third.

The importance of government payments to each group in the period covered relative to total income received may also be suggested. On the evidence of the same basic estimates, between 25 per cent and 30 per cent of the total income received by the lowest third over this period came from government payments, only slightly over 5 per cent of the total income of the middle third, and around 45 per cent of the total income of the highest third. These figures indicate the importance of actual money receipts from government at different income levels—they are definitely not to be taken as a measure of the total benefits received by

¹ Estimates of Charles Stauffacher, "The Effect of Governmental Expenditures and Tax Withdrawals upon Income Distribution, 1930-1939," in *Public Policy*, Yearbook of the Graduate School of Public Administration, Harvard University, 1941, pp. 232-261.

the various groups from the federal government. Allocation of benefits derived from government expenditures would be a far more complex problem than that faced in allocating money expenditures. Yet these estimates, when combined with federal revenue figures, may be of some value in indicating the redistribution of money income accomplished by the federal government's fiscal activities during the period covered. Since federal taxes over this period bore increasingly heavily on higher income levels and since spending was financed in considerable part by borrowing from the upper income groups, there appears to have been a clearcut redistribution of money income in favor of the lower income groups. However, state and local taxes bore heavily on the lower income groups, and when total tax burdens are compared with total money receipts from all governments, it is by no means certain how much such redistribution occurred. Detailed estimates of total tax burdens by income groups that emphasize this uncertainty are given in the following chapter.

CHAPTER 48

Public Revenues: Taxation

As WITH expenditures, it is convenient in discussing public revenues to segregate the case of relatively full employment from that involving cyclical fluctuations in employment and income or even "long-run underemployment." The segregation is somewhat unrealistic, but the discussion of public revenues under relatively full employment in this and the following chapter will be more realistically oriented in the more general chapter, "Fiscal Policy and the National Income," which considers the relations between "full" and "underemployment" for purposes of governmental fiscal policy.

Since this chapter and the next proceed largely on the assumption of relatively full employment, they are primarily concerned with taxation. With relatively full employment, there is ordinarily little need to resort to borrowing or the issuing of new money on a large scale to finance government spending, although a major exception to this statement arises in time of war. If borrowing or new money is used to a substantial extent in periods of full employment, the result is likely to be inflationary price increases, which are considered in the discussion of fiscal policy and the national income. Taxation has long been the common method of finance under ordinary circumstances.

Just as discussing expenditures without knowing the revenue sources used gives only a half view, so discussing revenues without the corresponding expenditures shows only part of the picture. But for simplicity, revenue sources are discussed in this manner; by fitting together the expenditure system (discussed in the preceding chapter) with the tax system (discussed in this and the next chapter) an overall view of the functioning of the public economy in "ordinary" times of relatively full employment can be obtained, and it is to this end that consideration of the various elements of expenditures and revenues is directed. In getting such an overall view, it is important to consider each of the major taxes separately; each must be evaluated to some extent as an individual levy. However, many errors have arisen through failure to view the tax system as a whole, and it is essential also always to bear in mind the overall picture in weighing the propriety of increasing or decreasing particular taxes.

Discussion of the "adequacy" of a tax system without reference to the

expenditures being financed has no more meaning than discussion of the "adequacy" of expenditures without reference to the means used to finance them. What are "adequate" taxes can properly be determined only by the sort of rough marginal comparison between tax sacrifices and expenditure benefits discussed in the preceding chapter. Discussion of the "adequacy" of any particular tax, except in a few special cases where the tax is used to finance special benefits for which direct payment is to be collected, is likely to mean even less. Thus for most purposes there is little point in discussing the "adequacy" of the income tax or of a retail sales tax, though there is point in discussing the adequacy of a special assessment levied specifically to cover the cost of a particular improvement that directly benefits the particular property on which the special assessment is imposed.

"Good" and "Bad" Taxes

Social criteria for taxes

Few subjects generate as much heat in popular discussion as taxes. Income taxes, sales taxes, property taxes, excess profits taxes, corporation taxes—all are staunchly advocated and bitterly attacked, often without very clear recognition of the criteria used in evaluation. Taxes, like anything else, can be evaluated only by reference to some standards of "good" and "bad," and it is important to recognize as far as possible what these criteria are, even though they may be quite different for different people. Five widely accepted criteria have been listed for use in discussing the private economy and restated in Chapter 46 for application to the public economy. Since the economist as such has no special right to set up broad social criteria, it appears best to continue to use these same five somewhat arbitrary "ends" as rough guides to evaluation with the same reservations previously observed.

With reference to the tax system these five criteria may be condensed into somewhat different and more precise terms. (1) Do taxes aid or hinder in the attainment of resource allocation consistent with consumers' preferences? In general, there is probably a presumption against interfering with the influence of free consumer choices on resource allocation, but there are numerous cases where tax policy may be designed specifically to interfere with this allocation—for example, taxes on liquor to discourage consumption, and the removal of taxes from forest land to discourage premature cutting. Closely related to this is the question, do taxes impede or encourage free choice of a trade or occupation? (2) How equitably is the revenue burden allocated? This is the specific criterion on the revenue side that is part of the more general criterion of equity in the distribution of income, and fundamentally reduces to the same issues involved in the broader question. Economists and other social scientists have spent much time and energy on the question of an equitable tax system, unfortunately often without facing the fundamental practical issue of income redistribution as such—the issue of the "haves"

versus the "have-nots." (3) What are the effects of taxes on investment, employment, income and business activity? This broad question covers both the usefulness of taxes in attaining greater stability and security from cyclical fluctuations and their attendant disruption, and the place of tax policy in the effort to attain a steadily higher plane of living for the nation. It covers both the more direct effects on employment and investment through costs and prices and the less tangible effects on public "confidence" and "uncertainty." Lastly, it covers the question of flexibility of adaptation to changing needs, both through the cycle and over longer periods.

"Equity" in taxation

In view of the basic importance of the issue and of the widespread attention given to it, special consideration is properly accorded the question of how taxes should be allocated. In the writings of economists two different basic criteria of "equity" in taxation have received strong support: (1) that the burden should be allocated according to "benefit received"; and (2) that it should be allocated according to people's "ability to pay." The second of these criteria appears in a variety of forms, all of which work toward a redistribution of income in favor of the lower income groups, although none of the formulations necessarily depends for its justification on the proposition that redistribution is itself the end sought. Some writers base their support of the various principles on general considerations of "equity" presumed to be more fundamental than income redistribution; others argue directly that redistribution is in itself the desideratum.

1. The "benefit" principle. The so-called benefit theory of tax allocation has a strong appeal for many persons. It seems to them equitable that anyone who receives a particular benefit from government should pay for that benefit. As a practical matter, this principle can be applied efficiently in only a few cases. How, for example, can the benefits derived from a battleship be allocated among the citizens of a nation? Who benefits most from maintenance of our system of courts and law enforcement? How is the gain to each individual from expenditures on forest conservation to be measured? Everywhere the attempt to base tax burdens on benefit received runs into such practical difficulties, which make precise application impossible and even a rough approximation often unattainable. Furthermore, in many cases adoption of the benefit principle would nullify the entire purpose of the expenditure being financed. This would be true, for example, in the case of relief payments, where the major aim is to provide additional funds for those unemployed.

On the other hand, there are some cases in which the benefit principle works reasonably well. Special assessments on real property to finance special benefits to the property, and gasoline taxes to finance improved highways are outstanding cases. However, even where the benefit theory can be applied, the problem of determining the benefits received by vari-

ous taxpayers is a treacherous one and the best that can be hoped for is a reasonable approximation to the benefit principle.

Insofar as the benefit principle is followed, the result closely approximates that obtained through direct sale of services by the government to consumers, except that the taxpayer is often forced to accept and pay for the benefit once the law is enacted to apply to the group of which he is a member. A further difference is that presumably there is no profit in the price (tax) charged by the government in providing the benefit. When viewed in this fashion it is clear that so long as the tax (price) just covers benefit received, no program of income redistribution is compatible with adherence to the benefit theory.

2. The "ability-to-pay" principle. Allocation of the revenue burden according to "ability to pay" has different connotations according to the interpretation given the phrase "ability to pay." If the criterion is the pragmatic one of "how many feathers you can pull without too much squawk or killing the chicken," the philosophic implications are quite different from using the criterion as a means of obtaining an equitable allocation of the revenue burden; yet these two attitudes may lead to similar results in practice. Leaving aside this attitude of obtaining revenue where the squawk will be least, advocates of the ability-to-pay principle usually take as the measure of ability net money income (with deductions for dependents), though total wealth has also often been suggested as an appropriate criterion. After the measure of ability to pay has been determined, the problem remains of determining the rate at which persons should be taxed, let us say on their incomes. Some argue that the rate should simply be *proportional*—that is, a certain fixed per cent of each person's income (for example, 1 per cent, giving a \$10 tax on \$1,000; \$100 on \$10,000; \$1,000 on \$100,000). Though the rate is proportional, this procedure of course takes more dollars from high than from low income groups. Others argue that rates should be *progressive*—that is, a higher percentage tax on high incomes than on low (for example, \$5 on \$1,000; \$100 on \$10,000; \$2,000 on \$100,000). This places a still heavier burden on the well-to-do than would taking a uniform proportion of their larger incomes or wealth. All argue against *regressive* taxation, in which a larger percentage of income is taken from the lower income groups (for example, \$20 on \$1,000; \$100 on \$10,000; \$500 on \$100,000).

It should be clear that this principle of ability to pay taxation has no exactness or "absolute" validity. Even when a fairly general agreement is reached on the "correct" measure of ability, there is no objective way of deciding whether rates should be proportional or progressive, and if progressive how steeply progressive. Clearly the principle provides no definite scale as basis for dividing revenue burdens. The "ability" phrase is, however, of some use and convenient for conversational purposes, especially since we seem to have reached a fairly substantial agreement that progressive taxation of incomes and inheritances represents the primary application of the ability-to-pay principle. When the term

is used in this book, therefore, it will be intended to mean progressively higher revenue burdens as the measure of ability to pay (net income) becomes larger.

Many writers have attempted to reach some more basic justification underlying the principle of allocating revenue burdens according to "ability to pay." One of the first suggestions was that each individual ought to have to make an "equal sacrifice." This proposition of course involves determining what constitutes an "equal sacrifice." It is commonly *assumed* that individuals in the same economic position (same income, same wealth, same dependents, and so on) derive the same satisfaction from spending one more dollar (that is, from a marginal dollar's expenditure). From this it would follow that individuals in the same economic position should bear the same tax burden if equal sacrifice is to be achieved. But when it comes to determining "equal sacrifices" for individuals in different economic positions, serious trouble begins. The assumption is commonly made that the more well-to-do an individual is, the less satisfaction he gives up by relinquishing a dollar. It seems sensible to say that \$1 paid in taxes out of an income of \$10,000 represents less sacrifice than \$1 paid out of an income of \$1,000. But there is no objective means of telling how much less satisfaction the marginal dollar represents to the richer individual. Since we can agree that \$1 generally means less to the rich man, it follows that he must pay more dollars if he is to bear an equal sacrifice with the poor man, but until the satisfaction of a dollar to each can be measured, how much more heavily the rich man should be taxed is indeterminate. Although it may be safe to assume that satisfactions between people can be compared sufficiently to indicate that rates should be progressive, the steepness of the progression remains arbitrary so far as the principle is concerned.

Lately numerous writers have suggested that the fundamental equity aim should be "minimizing total sacrifice" for society. If the assumption is followed that the marginal value of a dollar is lower the higher one's income, this principle logically leads to the proposition that the entire burden should be placed on the individual with the highest income until his income is leveled to equal that of the next highest individual, after which they should be taxed equally and exclusively until their incomes are decreased to the level of the next highest, and so on. No tax would be imposed on anyone as long as any other person still had a larger income. This conclusion follows from the assumption that in comparing any two income levels, a dollar given up by the lower income individual *always* means more sacrifice than a dollar given up by a higher income individual. Thus, if only two persons are considered, one with \$100,000 income and one with \$1,000 income, \$99,000 of the high income could presumably be taxed away before the loss of one dollar would entail as much sacrifice as the loss of the first dollar would to the \$1,000 income individual. "Minimum total sacrifice" would be obtained by taking each consecutive dollar where the sacrifice was least—that is, by allocating the tax burden in the manner indicated. How

rapidly this equalization would progress would depend on the amount collected and spent by the government.

On the other hand, many persons refuse to accept the assumption that the "satisfaction value" of a dollar is always lower the higher the income of the owner, pointing out that a very sharp reduction of the customary plane of living of a well-to-do person may well involve more "sacrifice" than a slight reduction by a somewhat lower-income individual. Whether or not this is so cannot be decided by the economist. But if the rigid assumption of the preceding paragraph is dropped, then one is thrown back on arbitrary interpersonal comparisons of "satisfactions" at different income levels, and the "minimum total sacrifice" principle is no more precise than its "equal sacrifice" neighbor in underlying support of the ability-to-pay position. It is important to recognize that all such "principles" of equity in taxation are likely to sound much more precise and "scientific" than they actually are. The basic question in formulating tax policy may be as much the direct one of how far redistribution of income should go as the one of what is the most "equitable" way of distributing the tax burden. Actually in modern thinking these two considerations are more and more closely interwoven.

Distribution of the Tax Burden

There are available certain rough quantitative estimates of the manner in which the total tax burden (federal, state, and local) has fallen on different income groups in recent years. Although the different studies have given roughly comparable results, it must be recognized that any such estimates are only approximations. This will be increasingly evident as particular taxes are considered, since it is often very difficult to estimate at all precisely where the burden of any particular tax does fall. Yet such estimates can be formulated more precisely than can estimates of the "burden" of inflation (which may arise in periods of relatively full employment when nontax means of financing are employed) such as were considered in general terms in an earlier chapter.¹

Through the latter part of the 1930's the per capita tax burden (federal, state, and local) ran slightly over \$100. Since then, this burden has risen tremendously, but inasmuch as the best estimates showing the distribution of the burden by income groups cover the late 1930's these figures will be considered first.² To make the estimates more meaningful

¹ Chapter 40.

² The estimates given here are based largely on Prof. Mabel Newcomer's analysis, which appears in *Studies in Current Tax Problems*, published in 1937 by the Twentieth Century Fund. Although these are still perhaps the best estimates available, they are open to the criticism of neglecting too much the possibilities of backward shifting of taxes through lower wages, rents, and prices of other productive agents. The meaning of this criticism will later become more apparent.

The other set of estimates generally recognized are those of Gerhard Colm and Helen Tarasov in *Who Pays the Taxes?* (Temporary National Economic Committee Monograph No. 3, 1941), which give similar results.

than they would be on an artificial per capita basis, they are calculated to indicate the average burden on "typical" families (of four or five persons each) at different income levels. But there are two general reasons why such estimates must be used only as rough indicators and only very carefully applied to specific cases: (1) we do not know how any particular family earns its income and how it spends it, both of which are critical for estimating its tax burden; and (2) we do not know at all precisely the final resting place of many taxes, even if we did have precise figures on the ways in which a family obtains and spends its income.

Typical family "A" with a \$1,000 income probably paid about \$200 in taxes (or around 20 per cent of its total income) in the period covered by these studies. The greater portion of this \$200 was probably "hidden" in the form of higher prices paid for taxed products, higher rents, and the like, or in the form of lower wages received by the wage earner. The family probably paid little in direct taxes, unless it owned its own home or unless sales and pay-roll taxes are called direct taxes. Probably well over half of the \$200 went to local governments, largely through the property tax (whether paid directly or through higher rents).

Typical family "B" with a \$2,000 income probably paid about \$350 taxes (or about 17 per cent of its income). Although this \$350 is a great deal more than the \$200 of the poorer family, it should be noted that family "B" paid a smaller percentage of its income in taxes than did family "A." The total tax system was thus "regressive" as between this and the lower income levels—the lower income groups paying a larger percentage of their income in taxes than did the higher. Family "B" escaped the federal income tax (and probably any state income tax), but was struck by most of the same taxes that hit family "A," especially the property tax. Here again, many of the taxes were probably "hidden."

Typical family "C" with an income of around \$5,000 paid probably \$1,100 in taxes (or about 21 per cent of its income). This percentage of income taken by taxes was only slightly higher than that for "A" and "B," although the absolute sum was much larger. Family "C" was subject to the federal income tax, although the payment was probably less than \$50, and it paid heavier commodity taxes to the federal government on many "luxury" items in the form of higher prices. Family "C" was also burdened by most of the taxes which fell on "A" and "B," with the exception of pay-roll taxes.

Typical family "D" with an income of about \$100,000 a year probably paid from \$45,000 to \$55,000 in taxes (or around 50 per cent of its income). For such well-to-do families, the federal income tax was the largest single item (amounting to possibly \$25,000), but almost all of the other taxes paid by lower income families hit "D" as well. Well over half of "D"'s taxes went to the federal government. Family "D" had a large "ability" to pay, and taxes fell heavily on it. A hypothetical family "E" with an income of \$1,000,000 might be considered to show that the percentage of income paid in taxes would have been much larger

yet; but when very high incomes are reached the tax paid depends so much on the sources of income and on the disposition of income that it is very difficult to make meaningful estimates. For example, income from many U. S. government bonds was tax-exempt, so a family receiving most of its income from this source would have largely escaped the income tax.

It is apparent that our tax system in the late 1930's to some extent conformed to the ability to pay and related "equity" principles in that it fell most heavily on the very rich, but it is equally apparent that the "middle classes" got off relatively very easy. The tax system was "regressive" with respect to income as between the low and middle income groups. To understand why this happened, it will be necessary to see the relative importance of the various taxes making up the American tax system and to analyze the impact of these various taxes on different income groups.

Since the beginning of the United States' large-scale rearmament program in 1940, and especially since our actual involvement in the war, federal taxes have risen unprecedentedly, though remaining still far short of paying the full cost of the war program. This increase has centered largely in three types of tax—personal income taxes, corporation income taxes, and sales and excise taxes. The first two are "ability to pay" taxes, especially the personal income tax at the steeply progressive rates now in force. Sales and excise taxes, however, generally fall heavily on the lower income groups and are likely to be regressive. Changes in state and local taxes have been relatively unimportant, though revenue from some state and local taxes has fallen off as civilian use of some goods (for example, gasoline for automobiles) has declined sharply.

In view of the rapidity of recent changes in the federal tax structure and of the rapidly rising and shifting nature of money incomes, it is very difficult to make any estimates of the changes that may have occurred from the tax distribution by income levels of the late 1930's. By the beginning of 1943, the overall tax rates in effect took over 25 per cent of the national income as compared with under 20 per cent in the late 1930's. Although total tax revenues almost tripled over the period, national income nearly doubled, so the increased tax payments by no means represented corresponding increases in "real burdens." People at any given income level paid far heavier taxes than during the 'thirties, but most people were moving up into higher income brackets. One other change is clear: where previously effective progression did not set in until annual income had reached around \$10,000, the reduction of income tax exemptions and the sharp increases in rates have pushed this point of rapidly rising progression much farther down the income scale, probably well under \$5,000. Up to early 1942, heavy excises apparently had raised lower income group tax burdens so much that the earlier regression between them and the "lower middle" income groups still existed, especially since income tax increases were effective only above the \$2,500 level for average families. The revenue bill passed in late 1942, however, relied

much more heavily upon income taxes, and effective rates on the lower middle income groups were substantially increased.

Shifting and Incidence of Taxation

Very often the person paying over the funds to the government does not actually bear the burden of the tax paid. A tax on the manufacture of cosmetics, for example, may be paid to the government by manufacturers, but be ultimately "shifted" forward to consumers in the form of higher prices or backward to cosmetics workers in the form of lower wages and to owners of other resources used in the cosmetics industry in the form of lower rents or lower capital values of their resources. The final "incidence" (or resting place) of a tax may be on the actual taxpayer or on consumers, workers, or owners of productive agents to whom it has been shifted. Obviously it is the incidence of the tax that is most important, rather than who pays the money over to the government. The difference between taxes, however, may lie even deeper than the question of who finally pays the tax. Some taxes may cause unemployment, or hinder investment, or reduce saving, or have other broad economic effects which go beyond the incidence of the particular tax. It is usually impossible to draw any sharp line between the incidence of a tax and its broader economic effects, as will become evident; but it is convenient to make a rough distinction of this sort for expository purposes. In a brief analysis such as this it is unfortunately impossible to go far in discussing the broad economic effects of various taxes.

It is safe to assume that a taxpayer will shift a tax whenever he can. The question is, therefore, when can a tax be shifted? A tax can be shifted *only* when as a result of the tax the taxpayer is able to obtain a higher price for the product or service he sells or is able to pay a lower price for the goods or services he buys. Hence, a price transaction of some sort is essential if shifting is to occur. If as a result of a tax the prices that a taxpayer receives or the prices that he pays are higher or lower *than they otherwise would have been*, the tax has been to that extent shifted. It should be emphasized that in discussing tax shifting it is common to assume *other things equal*, tracing through the effects of the tax as the only disturbance under discussion, just as the effect of an increased demand for some product was traced through, other things being equal, in analyzing the working of the price system. This simplification is necessary if one is to analyze the effects of any given tax—otherwise so many things are going on at once that we get nowhere. From this it should be clear that a mere rise in price following the imposition of a tax is in the real world not necessarily evidence that the tax has been shifted. The price rise may have come from some other cause operating simultaneously with the imposition of the tax. Empirical verification of tax shifting is thus virtually impossible in many cases, since it is so difficult to isolate one cause and its effects in the multitude of forces simultaneously at work in most parts of economic life.

A tax can be shifted only by making a price different from what it

would have been without the tax. Since a price will ordinarily change only when demand or supply, or both, change, if a tax is to be shifted it must affect either the demand for, or the supply of, the good or service in question. For example if a tax on a manufacturer is to be shifted forward to consumers, it must either increase demand for the product or decrease the supply of the product. There is no particular reason to suppose that demand will be increased as a result of the tax, so the tax can be shifted to buyers of the good only if supply is reduced. This may well happen, since the tax raises the cost of producing the article in question, and higher costs, demand remaining unchanged, will tend to restrict output in both the short and long runs. Output will be restricted somewhat in the short run if marginal costs are increased by the tax, since producers will be led to reduce output in order to maximize profits or minimize losses. Output will be restricted further in the long run, since profits will now be smaller and firms will gradually tend to drop out of the industry, moving to other industries where profit opportunities are better. In general, any tax that falls differentially on various investment opportunities will tend to be at least partly shifted to consumers since investment will move from the more heavily taxed industry to less heavily taxed industries, thereby reducing the supply of the specially taxed product and causing the tax to be shifted. This same reduction of output means decreased demand for production services in the taxed industry, thereby also leading to backward shifting.³

This oversimple example shows the general sort of analysis that must be applied to discover whether or not taxes are shifted. Some taxes are almost certainly shifted; others (for example, the personal income tax) have little direct effect on either supply or demand for any product or service and hence are unlikely to be shifted. The limiting assumption of "other things equal" is virtually essential, especially for elementary analysis, if anything like precise results are to be obtained. In addition, we shall assume throughout, except where the contrary is specified, a constant total volume of money expenditures per time period, and for the present avoid thereby the complications that arise when the possibilities of offsetting tax effects by monetary and fiscal policy are considered. These simplifying assumptions are well justified by the nature of the problem if care is taken that the conclusions are used only in the light of the assumptions, and if the assumptions are dropped in certain cases where they are counter to the nature of the problem.

There are few economic problems as complex as the determination of the incidence and economic effects of any given tax. However, there are certain reasonably simple conclusions that can be drawn about the incidence and effects of most of our important taxes, and it is these conclusions in which we shall be primarily interested. Such analysis is of

³ This introductory statement is intended only to give a general idea of the pattern of price analysis that is useful in discussing various taxes in the following sections, and is strictly accurate only on the assumption of pure competition although it is essentially correct in application to most other type-situations as well.

necessity considerably oversimplified, and the conclusions drawn therefore only tentative and general. But the analysis is of considerable value, both for showing the nature of the most important effects of our various taxes and as an exercise in applied economic principles. Unless we know at least roughly where the burden of any tax ultimately falls, there is little sense in talking about its place in an equitable tax system.

Excises

Excises are taxes on particular commodities assessed when the commodities are produced or when they are sold. Excises are usually either so much per unit sold (as so much per pack of cigarets) or so much per dollar's worth of product (as so much per dollar's worth of cosmetics). Few persons realize the range of products covered by such taxes. The federal government alone, in its regular reports on internal revenue collections,⁴ lists excises on 52 types of commodities and services, most of which people seldom think of as subject to special taxation. We all realize that we pay special federal taxes on tobacco, alcoholic beverages, admissions to theatres, and long-distance telephone calls. But few of us realize that special federal taxes have for years also been imposed upon mechanical refrigerators, electrical energy, club dues and initiation fees, radio and phonograph sets, automobiles, automobile parts and tires, motor fuels of all sorts, and toilet preparations, each of which yielded over \$5,000,000 in 1941, plus a large number of others of varying fiscal importance. On top of these federal excises, which are collected from manufacturers and wholesalers, are imposed state and local excises, all in addition to the retail sales taxes that became so common during the 1930's.

Because people remain unaware of the existence of many excises, such levies are relatively easy to impose. They mean plucking the feathers where the squawk will be least. Naturally, in the political processes of tax-making, levies that are relatively invisible to the voters make a strong appeal to legislators, and such taxes are often enacted instead of others generally regarded as "better" from all points of view except that of minimizing the squawk.

Incidence

It is the common impression that excises are passed on to consumers through higher prices. There is considerable truth in this statement but it does not give a completely accurate picture. Such taxes are usually borne in part by at least some of the productive agents employed in the industry as well as by consumers. The incidence of an excise depends largely on whether the industry is partially monopolized or purely competitive, and on the relative elasticity of demand for the product and the elasticities of supply of productive agents to the taxed industry. Let us take an example of a tax per unit of output (as per pack of playing cards) to analyze the probable incidence of such excises. The analysis

⁴ *Bulletin of the Treasury Department*. Figures used are from February 1942 issue.

will be divided for convenience into short-run and long-run incidence under competition, and then under simple monopoly and other monopolistic arrangements, with consideration of the broader economic effects postponed temporarily.⁵ In order to provide the framework for later incidence discussions for other taxes, the excise case is discussed in considerable detail.

1. **Pure competition.** Suppose a new excise is imposed on playing cards, which we shall assume are produced by a competitive industry.

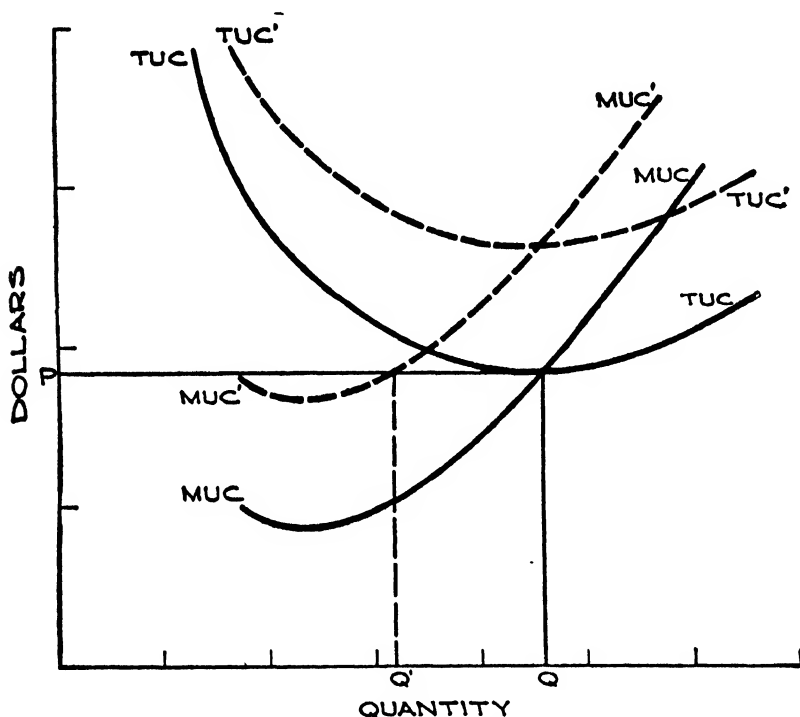


Fig. 48—1. Effect of excise tax on purely competitive seller.

What will be the immediate (*short-run*) effects? The first effect is, of course, that the tax raises the cost per pack of playing cards by the amount of the tax—both the average and marginal unit cost curves are higher throughout by the amount of the tax, since for each additional pack produced the firm has to pay the excise tax. If an original position of long-run equilibrium is assumed, the shift may be shown very simply, as in Figure 48—1. Before imposition of the tax, the price is P , just high enough to cover the lowest total unit cost of production, with each firm producing OQ units in long-run equilibrium. Curves TUC' and MUC' then represent the cost curves after the imposition of the tax.

⁵ Analysis of the incidence of an ad valorem excise would differ slightly from that of a per unit tax, since with the former the effect of the tax on the entrepreneur is to shift his revenue curve to the left rather than to raise unit production costs.

Since profits are maximized by carrying production up to the point where marginal unit cost equals marginal unit revenue (here price), after the tax-induced increase in marginal costs, in order to minimize losses each firm will produce only OQ' units instead of OQ , equating marginal unit cost and marginal unit revenue at that output. The first effect of the tax will be a reduction of output by all firms.

When all firms reduce output somewhat, two results are likely: First, price will rise, and second, fewer workers and other productive agents will be demanded. If factor prices are flexible, they will be forced down somewhat. We might proceed to analyze in detail just how much higher the price would be and how much lower the prices of various productive services, but this is better postponed to the long-run analysis; the principles are the same in both cases. It is safe to say that in the short run, incidence of the tax is likely to be partly on consumers (through higher prices), partly on workers and owners of variable productive agents (through lower wages and raw material prices), and partly on those receiving payments for fixed agents and interest (the entrepreneur-owner being in this group, since even with reduced output a loss is being made—price remains lower than the new unit costs). The entire burden of the tax will be distributed over these three groups.

What will be the *long-run* incidence of our excise? The tax will necessarily be shifted in the long run, either forward or backward. The immediate effect is to push costs above prices, leading to losses in the playing-card industry. When losses are being suffered in any industry, firms move out and investment shifts to other more profitable opportunities. And since our excise is a special tax on the playing-card industry alone, other investment opportunities are now relatively more profitable than playing cards. As fixed agents wear out in producing cards, they are not replaced; instead the depreciation funds are used for investment in other industries.

This movement of firms out of card production will continue *until the price of playing cards has again risen to cover average total unit cost at the lowest point on the average unit cost curve*—that is, until investment elsewhere is no longer relatively more attractive than in the making of playing cards. When price has risen and/or costs fallen to this extent, a new long-run equilibrium is attained, with price again equal to lowest average cost and fewer firms in the industry. The tax has been completely shifted from the manufacturers forward to consumers or back to owners of productive services. In the long run, the tax cannot remain on producers since as long as investment in playing cards is less attractive than elsewhere (which it will be until price again covers all costs, including the tax), firms will drop out of the playing-card industry.

But this does not say in which direction the burden will be shifted—forward to consumers or backward to laborers and owners of relatively specialized productive agents. In this particular case it is likely that consumers will ultimately bear most of the burden. Why?

As firms move out of the playing-card industry, fewer cards will be

produced and the demand for productive services will decrease. The question then is, will it be easier for the tax to be passed on in higher prices or back through lower wages, rents and material prices? Let us assume, probably realistically, that the demand for playing cards is relatively inelastic—that is, that purchases of playing cards would drop off only slightly as prices rose. On the other hand, if card manufacturers offer lower prices for paper stock or for common labor (that is, for unspecialized resources that can easily move from one industry to another), they will probably be unable to obtain any paper stock or labor. The playing-card industry must pay going prices for such productive agents and raw materials or not get any; in technical terms, the elasticity of supply of these agents to the playing-card industry may well be almost infinite. This is especially likely to be true since the industry constitutes so small a portion of the total demand for these agents.

Under these assumed conditions of inelastic product demand and highly elastic factor supplies, it is obviously much easier for the price of cards to rise than for costs to be pushed down. In fact, on the assumption of completely unspecialized agents, factor costs to the card industry would remain just as high as before in spite of its decreased demand for labor and paper. Under these circumstances, firms will move out of the card industry until price has finally risen by the full amount of the tax per pack, shifting the tax completely to consumers. The critical elements in determining the direction of shifting are the relative elasticities of product demand and factor supplies.

For most excises under pure competition forward shifting is probably the general situation. But the assumption made above that all resources used are completely unspecialized is not always true, even in the long run. To illustrate, suppose the playing-card industry takes almost all the output of some grade of paper pulp that is not useful in any other industry. In that case, the supply of the paper pulp to the card industry would be very inelastic—even if the card industry offered a much lower price it could probably still get almost as much pulp, since there is no place else for the pulp to be sold. The tax might be in large part shifted backward in lower pulp prices. The fewer the alternative opportunities for productive agents (that is, the more specialized they are), the more such agents are forced to take lower payments as a result of decreased demand by the card makers; the more alternative opportunities there are for productive agents (that is, the more unspecialized they are), the less they are forced to take lower payments as a result of decreased demand by the card makers.

This analysis of excise shifting and incidence under competition may be summarized briefly as follows: A special excise will be completely shifted, because investment will move away from the taxed industry to other less heavily taxed investment opportunities until price again covers costs in the taxed industry. The direction in which the tax is shifted will depend on the elasticity of the demand for the taxed product relative to the elasticities of supply of the productive agents used; shifting will be largely

in the direction where the elasticity is lower. In the long run, most resources are unspecialized and will tend to move away from the taxed industry if it offers lower than going wages and other factor prices, so generally taxes cannot be shifted backward and are largely passed on to consumers. Further, in practice monopoly groups are strong among resource sellers, which permits an even stronger resistance to backward shifting as compared with the assumed competitive card market. But if some particular productive agents are relatively specialized even in the long run, a portion of the tax will be shifted back to them instead of

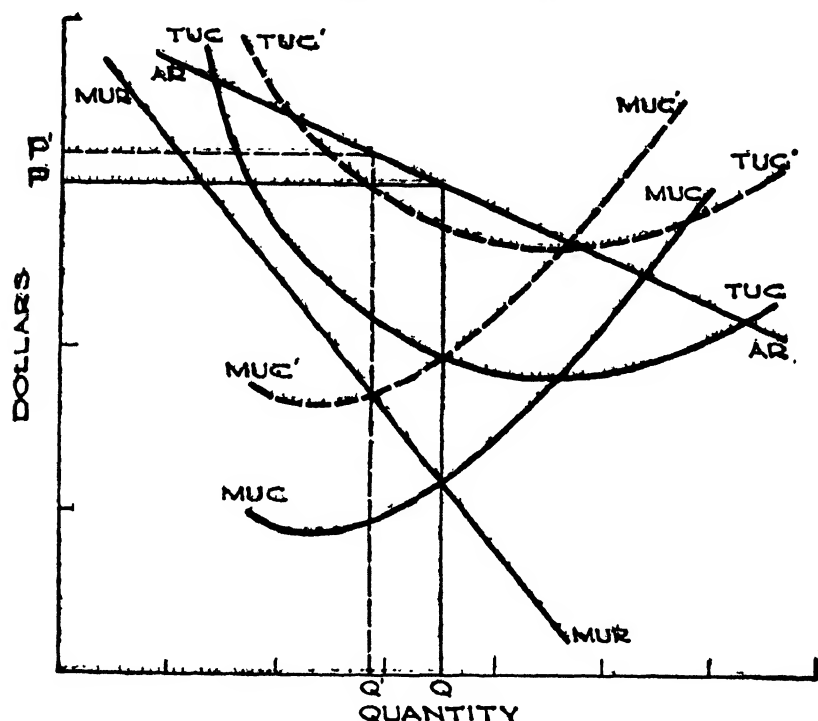


Fig. 48—2. Effect of excise tax on monopolistic seller. *

forward to consumers, unless monopolistic powers of resource owners prevent this development. The short-run incidence of the tax is much more uncertain. Temporarily, in view of short-period immobility and specialization, entrepreneurs may well be able to pass back the tax through lower rents, raw material prices, and possibly, though not likely, wages; but gradually productive agents will move to other industries, forcing costs back up in the taxed industry if it is to get any resources.

2. Simple monopoly. Analysis of the incidence of an excise imposed on a monopolized product where restriction against entry of new firms is maintained is similar to that of the competitive case, but the results are somewhat different. Let us again assume an equilibrium position, with the monopolist maximizing profits by producing up to the point where marginal unit cost equals marginal unit revenue and charging the price shown for that output by his average revenue curve. Figure 48—2 shows

this case, with the monopolist producing OQ units before the tax and charging price OP . As in the competitive case, imposition of the excise tax raises both marginal and total unit costs by the amount of the tax. After the tax, to maximize profits the monopolist will reduce output to the new intersection between marginal unit cost and marginal unit revenue (OQ'), and will raise price to the corresponding figure on his average revenue curve (OP'). The short-run effect of the tax is thus a reduction of output and a higher price, just as in the competitive case. And as before, part of the tax may be shifted back to productive agents used in the industry since with reduced output there is less demand for these resources. How high the price is immediately raised and how much of the tax is temporarily passed back to productive agents depends, again, on the elasticity of demand for the product, on the degree of specialization and mobility of the resources used, and on the monopoly powers of resource owners, but in this case also in considerable part on the anticipations and discretion of the monopolist card producer. Shifting, again, will occur primarily in the direction where the elasticity is the lowest—that is, where the resistance to shifting is least. And the monopolist probably will have, at least in the short run, to absorb part of the tax through lower profits.

What will be the long-run incidence under these circumstances? There is one basic difference between this and the competitive case. Under pure competition the tax will be completely shifted, largely by the movement of investment out of the taxed industry. But here the monopolist is likely to be still making a profit even after the tax, even though the tax reduces his earnings. Therefore, any substantial long-run shift of investment out of the monopolized industry is unlikely. The monopolist is still better off here than elsewhere, though of course he may reduce his scale of plant, and the long-run incidence may be similar to that in the short run. In the long run, however, the possibility of backward shifting to specialized, immobile productive agents is much less than in the short period.

3. Oligopoly and monopolistic competition. The incidence of an excise on a product produced and sold under oligopoly or monopolistic competition is very uncertain, since the whole pricing process is unpredictable and subject to "cooperative" agreements. It is impossible in such cases to illustrate at all precisely what price and output will be after the tax as compared with before it. In general, it is safe to say that the tax will be largely shifted either backward or forward in all cases where free entrance to the cartel or price-leadership arrangement exists or where close duplication of differentiated products is possible. Shifting is then likely because profits tend to be eliminated in such cases (through higher costs from overinvestment in the industry rather than through lower prices) whether or not there are special taxes, and hence the special tax will probably induce shift of investment away from the taxed industry.⁶

⁶ Excises on products produced in oligopolistic industries may under certain circumstances be met almost entirely through increased production efficiency. If entry to such an industry is free, "too many" firms would be in the industry, each operating

If free entrance does not exist and the oligopolists continue to make profits in spite of the tax, the case is closely analogous to that of simple monopoly, and part though not all of the tax is likely to be shifted. How much shifting occurs in such semicompetitive cases therefore depends largely on the degree of monopoly power exercised by the various buyers and sellers involved. But whatever the restrictive power exercised by a seller over production, his possibilities of shifting taxes ultimately depend on the elasticities of supply of productive factors and of demand for his product. When, as is frequently the case, a tax is imposed on a partially monopolistic producer who deals in turn with partially monopolistic sellers of productive services and buyers of finished products, formal price analysis retains little usefulness in determining tax incidence, and the result is largely determined by relative bargaining positions and institutional conditions. For example, the incidence of a tax on the production of steel parts for construction purposes would be highly indeterminate, since producers are monopolized, sellers of productive services and materials in the many parts of industry are well organized, and many users of such steel are very large-scale buyers, in turn possessing considerable monopoly power in their own selling markets.

Even in the simpler cases, the process of shifting may move very slowly and jerkily. Monopolists especially may not set output and price in the purely "rational" manner which is often attributed to them. General business conditions play a great part in determining the ease and speed of shifting—taxes are always much easier to shift in good times than in bad. Even with predictable conditions in the taxed firm, great imperfections may exist in the market for productive services that impede or facilitate backward shifting. Producers may have difficulty in estimating elasticity of both product demand and supplies of productive agents. In the real world, conditions are much less precise than in textbook analysis. Yet the central conclusions reached in the preceding sections probably hold good through the complex of interacting forces at work in any real-world market.

Evaluation

1. Effects on production and consumption. An excise clearly discourages the allocation of productive resources to the taxed industry. By the same token it discourages consumption of the taxed product. If the incidence of excises is primarily on consumers through higher prices and reduced consumption, how good such taxes are, judged by the ability-to-pay standard, depends largely on the income groups represented by consumers of the taxed products. A tax on "necessities" is generally agreed to be regressive; a tax on "luxuries" is commonly thought to

short of the optimum scale. As some firms withdraw owing to losses resulting from the tax, each of the remaining firms *may* produce more than before, at a lower cost per unit of output. In this case the tax *could* be paid out of savings from more efficient operation without either reducing the incomes of productive agents in the industry or raising prices to consumers.

be progressive. These loose generalizations depend, however, on how "necessities" and "luxuries" are defined. If "necessities" are those goods that everyone (or at least the lower income group) consumes in considerable amount, then taxes on such goods are clearly regressive, since a larger proportion of total income will be spent on them at the lower than at the upper income levels. If "luxuries" are those goods that people do not "need" in order to live, such as fur coats, cigarets, Cadillacs, and liquor, taxes on them may or may not be progressive. The standard important examples of "luxury taxes" are those on liquor and tobacco, which are intended both to raise revenue and check consumption of these "luxuries," but in fact the demand for both these products by rich and by poor is relatively inelastic. The tax of about 8 cents per pack on cigarets in many states often leads not to lower expenditures on cigarets but instead to lower expenditures on clothing and food, since many persons give up what are usually called "necessities" rather than cigarets. The same is true of liquor, and to a lesser extent of numerous other similar commodities. "Necessities" can be defined in any way one likes, but definitions based on moral standards should not blind us to the inelastic nature of the demand for many products labeled "luxuries" nor to the real effect of taxes on them. To be sure, cigaret and liquor taxes do discourage the consumption of these products, especially by the poor. But unless care is taken, taxes on "luxuries" are likely to be highly regressive, and unfortunately, many existing excises are on commodities heavily consumed by the lower income groups. If excises are imposed on such goods as mink coats and caviar, grounds for criticism of them as regressive vanish, but taxes on such articles can yield little revenue since so few purchases of them are made. If excises are to yield large sums they must fall on goods heavily consumed by the middle and lower income groups.

Taxes directly intended to restrict consumption are often called "sumptuary" or "social control" taxes. It should be clear that all excises, however intended, have some "sumptuary" results—higher prices restrict consumption of the taxed good or some other good or service, whether this effect is intended or not. The common argument is that sumptuary taxes are of double benefit, that they both control consumption and yield revenue. But to the extent that a tax is a good sumptuary tax (that is, limits consumption), it does not yield revenue. There is a direct conflict between the two aims. A very similar purpose may be fulfilled by excises in wartime when it becomes desirable to restrict consumption of particular commodities that compete directly for resources needed in war production. Especially where demand is elastic, such taxes may prove to be very useful. This device was used to a limited extent in the United States during the recent rearmament period, but more stringent and direct controls through materials priorities and allocations were rapidly imposed after the declaration of war in 1941. As with peacetime sumptuary taxes, much confusion has existed as to the possibility of increasing yields and restricting consumption simultaneously by imposition of such excises.

2. Effects on the level of employment and national income. Any tax in itself is deflationary since it draws funds out of the income stream. When the funds are respent by the government they may more or less than offset this deflationary effect. The net effect of the government's fiscal policy depends on both the method of raising the funds and the method of spending them. Therefore, in analyzing the effect of different taxes on the level of employment and national income, it is useful to make given assumptions as to the manner in which the funds obtained are to be spent and simply to compare the effects of different taxes, other things being equal. Such a simplifying assumption permits us to distinguish fairly sharply between the effects of different levies. Hence, except where otherwise specified, the assumption will be made that tax receipts are promptly spent by the government in the same pattern of expenditures as has previously obtained.

The immediate effect of a new excise, since it falls directly on marginal costs, is to reduce output and employment in the taxed industry. If the tax is imposed in good times, and especially if it falls on only a narrow sector of the economy, this unemployment is likely to be eliminated quickly through absorption of the displaced workers and other resources in other industries. Product demand is likewise less likely to be highly elastic in prosperous periods. Excises are more deflationary than other taxes that do not fall directly on business costs, but under prosperous conditions they alone are not likely to be a critically deflationary factor. In depression periods, however—the condition under which many current excise and sales taxes were imposed—the strongly deflationary effect of new excises is sure to be much more disruptive. Workers unemployed because of such taxes find it virtually impossible to get other jobs at acceptable wages. Increased unemployment intensifies the spiral contraction of money incomes and expenditures. Barriers to mobility, such as lack of information, inability to obtain needed retraining, the high cost of shifting jobs and locations, and union restrictions on entrance, bulk especially large in depression. If the excises apply to many products instead of only a few, the difficulty of re-employment is correspondingly intensified.

It might be thought that workers and other resource owners would rather take lower prices for their services than accept unemployment as a result of deflationary taxes. In fact, however, wage rates especially have proved highly rigid against downward pressure, except in nonunionized industries. This downward wage rigidity removes much of the possibility that selling prices and sales of the taxed products could remain roughly constant while costs fall to absorb the tax. But even though costs should prove flexible and absorb the tax burden, resource owners' incomes would thereby be reduced and deflationary pressure elsewhere would result. Excises such as we have, however their incidence develops, exert a strong and direct deflationary pressure that is equaled by few, if any, other taxes. Pay-roll and sales taxes are the only important close competitors. When compared with, say, progressive income taxes that do not fall on business

costs or draw so heavily on funds that would have been spent anyhow, the depressive effect of excises becomes especially plain.

The major arguments for excises are (1) that they raise large sums rapidly if applied to widely consumed commodities (which therefore necessarily are those goods consumed by the lower income groups where consumption is heavily concentrated), and (2) that, being "hidden" taxes, they arouse relatively little popular resentment. If one wishes to disregard equity considerations, widespread heavy excises may be an effective anti-inflationary tax source in boom periods. However, in depression this major "virtue" of excises in raising large sums with a minimum delay becomes a "vice" since these very taxes so directly increase the unemployment that the funds are raised to combat. Lastly, except as political concessions, it is hard to justify hidden taxes in a democracy, where "tax consciousness" is essential to economical administration and to genuine popular interest in government.

Motor fuel taxes

The excise on gasoline (more broadly, the taxes on motor fuels) levied by the federal and all state governments deserves special attention because of the special manner in which most such revenues are spent. Since receipts from these taxes are used primarily in improving the highways and streets, the taxes can be justified to a considerable extent as benefit levies, roughly comparable to the financing of special real estate improvements by special assessments. In general, like other special excises the taxes are borne by the consumer (motorist) in the form of higher prices for motor fuels, although there may be some backward shifting to resources used in producing the fuels and perhaps a minor amount of forward shifting to consumers of products sold by business firms using gasoline in salesmen's cars, and so on. Insofar as the motorist pays the tax, it may be regarded as a sort of substitute for a price charge for the use of roads. The purchaser of gasoline is in fact buying, through the gasoline tax, mileage on government roads. A high yield from a gasoline tax indicates the willingness of road travelers to pay the tax levied to keep up the roads. Moreover, with improved roads the demand for gasoline may in turn be increased. In this way, the sumptuary effects of the tax are partly avoided.

Although motor fuel taxes may be largely justified as benefit levies, there are difficult problems involved. For one thing, financing the roads by present compulsory taxes on all motorists is said by many to give a subsidy to trucks and buses using the roads, while railroads are forced not only to maintain their own roadways but even to pay taxes on them. While some types of trucks and buses are now so heavily taxed that they apparently do bear their fair share of maintaining the roads, many others are grossly undertaxed. For another, the gasoline tax is obviously not a complete means of reaching those who benefit from highway improvements; many persons who buy comparatively little gasoline benefit greatly from good roads and streets. For example, a farmer is benefited by a good

road past his property, even though he may drive to town only when he harvests and delivers his crop. A gas station owner is benefited by the improvement of the road on which he is located, even if he himself should never drive a car nor ride in one. Therefore it seems equitable that roads and streets should be financed partly by owners of property benefited by the roads and streets; and this is commonly done by using some of the property tax revenue for roads and streets. Also, most states require vehicle owners to purchase licenses, using the funds to augment gasoline tax revenues. To the extent that gas tax revenues do not result in corresponding expenditure on highway finance, obviously the gas tax must be judged by the same criteria as other excises.⁷

Retail Sales Taxes

The primary difference between retail (often called "general") sales taxes and excises is that excises apply only to selected commodities while the sales tax applies to all, or a very large number of, commodities at a common rate of so much per dollar of sales. When, owing to the increased need for relief payments in the face of shrinking property tax revenues, the states found themselves badly pressed for funds at the depth of the depression, over half of them turned to some sort of sales tax as the solution. The primary virtue claimed for the sales tax was that it would raise money and raise it quickly. In addition some city sales taxes were imposed, but for the most part the levy has been a state tax. At the outset of World War II, state sales taxes produced about \$750,000,000 annually, and for many states they had completely or largely replaced the state general property tax. More recently the general sales tax as a federal levy has been pushed to the fore as a method of financing the war program and holding down consumer buying power.

Incidence

The incidence of a sales tax depends partially on whether the tax is "general" (nationwide) or covers only a certain economic area, such as a state or city. Most of the sales taxes enacted during the past decade to provide depression-period revenue were state taxes. Although proposals were repeatedly made for a federal sales tax at the retail or wholesale level, these were consistently opposed by the Roosevelt Administration, and it was not until the great pressure for funds in the war emergency that a federal sales tax became politically expedient. Obviously the distinction between a general sales tax and a very wide range of excises is hazy. If excises are placed on the bulk of individual commodities generally consumed, the effect is almost identical with that of a general sales levy.

It is commonly said that sales taxes are passed on to consumers. A sales

⁷ This does not mean that the actual gas tax and auto license funds must be earmarked for highway expenditure to be justified as a benefit levy, but simply that a corresponding expenditure be made on highways. Earmarking of particular funds for particular uses is apt to lead to disastrous inflexibility in governmental expenditures.

tax in a state surrounded by states that do not have similar taxes may be especially difficult to shift forward, however, even though the state law commonly provides that the tax shall be added to the prices of commodities sold. Suppose Iowa has a 2 per cent sales tax and Minnesota has none, so that the effective cost of articles to consumers is 2 per cent higher in Iowa. This will help induce Iowa residents, let us say, near the northern border, to purchase in Minnesota and residents elsewhere in the state to buy by mail if convenient. To the extent that consumers are "mobile" as between states, Iowa merchants must either reduce list prices to offset the tax or lose business. In this case the tax would probably be partially shifted back to sellers of productive agents in Iowa, insofar as they were relatively immobile and inelastic in supply, and it would be partially borne at least temporarily by store operators, especially where some monopolistic profits existed. To protect against this danger to entrepreneurs and workers of the taxing state, "use taxes" have been widely enacted imposing a tax on all goods brought into the state from states having no corresponding sales tax. This prevents residents from avoiding the tax by buying outside the state, and helps retailers shift the tax to consumers.

With the rapid spread of state sales taxes, accompanied by use taxes in many cases, the practical possibilities of avoiding taxation by shopping out of state decreased rapidly, so that even aside from new federal consumption taxes the effect has been similar to that of a general sales tax. Under these circumstances, whether the law requires that the tax be listed separately to the consumer or simply charges the retailer (expecting that the burden will be shifted forward), the incidence may be on consumers, it may be forced upon owners of productive agents, or it may be partially borne by entrepreneurs.

Given any assumed money incomes, consumers can buy less at higher prices than at lower, so that if the tax is reflected in higher prices the amount bought declines. Such a decline reduces the demand for productive services, leading to partial backward shifting through lower wages, rents, and materials prices. Whether the sales tax on any particular commodity is shifted forward or backward or remains partially on the businessmen who pay the tax to the government depends on the relative elasticities of demand for the commodity and of supply of the productive services used, and on the monopolistic or competitive nature of the markets involved. Therefore, the incidence of a general sales tax may vary considerably as between different commodities. In a sense, it is similar to a large number of specific excises, though the more general nature of the sales levy makes forward shifting through reallocation of resources away from taxed industries much less easy.

The one generalization that may be made about the incidence of a general sales tax is this: it will decrease the incomes of productive agents relative to the prices of finished commodities. This proposition holds good whether the tax is retail or wholesale, state, federal, or local. Since a general sales tax covers almost all commodities widely consumed and therefore affects almost all persons as sellers of productive services, on

the broadest level it makes little difference whether technically the burden is passed forward through higher prices or backward through lower money incomes since the people involved are the same either way. However, the differential effects on different consumers and owners of productive agents are likely to be great, depending on the direction of shifting. Therefore, a more detailed analysis of the incidence by particular commodities or commodity groups is required if a thorough picture is to be obtained for any practical purpose. To some extent the burden may remain on entrepreneurs, but the chances of this as a long-run condition are very small except where restriction of entry protects monopoly profits.

Evaluation

1. Effects on consumption and the distribution of income. If a sales tax covers all finished goods, its differential effects on the allocation of resources are largely random, since it applies at a flat rate to all goods sold. Exemption of housing and sales of services may lead to a shift of resources to these fields, but property taxes are likely to be an even more serious deterrent to new housing than are sales taxes to production of taxed goods. When broad classes of goods, such as food products, are exempted from a sales tax, a substantial shift of resources into the tax-free area may result.

Sales taxes are seriously regressive, especially general sales taxes that fall on all consumers' goods. A family whose income is \$1,000 is likely to spend almost the entire amount except for rent payments on taxed consumption and (assuming a forward shifted tax of 5 per cent) pay a tax of \$50, disregarding rent. But a family whose income is \$10,000 will spend only a portion of this amount on consumption goods (let us say, \$5,000) and invest or hold the rest. The wealthy family is thus taxed 5 per cent on only \$5,000, making a total sales tax burden of \$250, or only 2½ per cent of total income. Even though the wealthier family pays more dollars, it pays a smaller per cent of its income. Exemption of such groups of "necessities" as food mitigates but does not eliminate this regressive effect. Only by exempting such broad groups of commodities as to convert the tax in effect into a group of special "luxury excises" can the regressive effect be completely eliminated, and such a conversion, by removing the widely consumed goods from the taxed list, reduces the yields so drastically as to make such a sales tax fiscally little more desirable than a few particular excises alone would be.

2. Effects on the level of national income and employment. Sales taxes imposed in depression periods are not only regressive but are also among the most deflationary of all taxes. Such levies are likely to be especially depressive at such times because they either directly increase sellers' marginal costs (if not shifted) or reduce consumers' buying power (if shifted either forward through higher prices or backward through lower incomes). Since they fall especially heavily on the lower income groups, who spend most of their incomes on consumption, sales taxes are relatively unlikely to draw on otherwise idle funds, as contrasted, for

example, with a progressive personal income tax. In depression periods little economic justification can be found for enactment of such levies. We could probably have found few worse taxes to finance government activities in the long depression of the 'thirties than the sales and excise taxes widely adopted. Only pay-roll taxes on the employer, which were also popular, appear to have a more directly deterrent effect on employment.

By the same token that a general sales tax is especially repressive in a depression, it is a strong anti-inflation weapon in expansionary boom periods, though this effect obtains largely because of its regressive nature in drawing heavily on consumption expenditures of the lower income groups. Progressive taxes, on the other hand, are far more likely to be paid partly out of funds otherwise saved, and are therefore less effective anti-inflation devices when the object is to reduce consumer buying power. Whether this anti-inflation advantage of general sales taxes outweighs their highly regressive nature in inflationary periods is a question of judgment. Some writers feel that even a considerable degree of inflation would be less inequitable than a heavy general sales tax. Others feel that even regressive taxation is far preferable to the inequitable, disruptive spiral of inflation.

Pay-roll Taxes

With the vast program of "Social Security" benefits enacted by the "New Deal" in the business depression of the 'thirties, pay-roll taxes sprang into a position of major importance, yielding somewhere around a billion and a half dollars annually. The taxes were inaugurated almost exclusively to finance unemployment and old-age insurance plans. In the case of unemployment insurance, every employer (except those specifically exempted by law) must pay into the insurance fund a sum equal (in 1942) to 3 per cent of each employee's wage. In the case of old-age insurance the employer must also pay into the insurance fund a certain per cent (1 per cent in 1942) of employees' wages; and a like sum is also deducted by the employer for the government from the wage paid out to the employee, making total contributions of 2 per cent per worker.

The Social Security Program and some effects of financing it through pay-roll taxes will be discussed further in Chapter 51. The present section presents only a brief analysis of the incidence and economic effects of the taxes themselves.

Incidence

The incidence of both parts of the pay-roll tax (the part deducted from wages and the part paid by employers) is primarily on the workers covered. While this general conclusion appears well supported, the widespread imperfections in the labor market, coupled with legal and extra-legal controls over wages, make the process of shifting highly uncertain and jerky in particular cases. In cases of wage inflexibility, moreover, the incidence may well fall in the form of unemployment for the workers involved. It is convenient to divide this brief discussion

into separate parts dealing with the tax collected out of wages and that imposed on the employer, respectively.

1. The incidence of the portion of the tax deducted directly from the employee's wage is almost invariably on the employee. There is no one onto whom he can shift it backwards, and it is very difficult to shift it forward to employers and perhaps hence to consumers via higher wages, since there is nothing in the tax to increase the demand for labor. Wages, therefore, can be pushed up in the occupations covered only by a withdrawal of workers or by forcing the hand of monopsonistic employers to pay wages higher than before for the same number of workers. Withdrawal of workers from covered occupations to untaxed jobs is unlikely because there are very few occupations that are uncovered. Agriculture and domestic service are the only two important cases (except that establishments with less than eight employees are exempt from the unemployment insurance tax). By moving to these occupations, workers would escape paying the tax; but they would also lose both unemployment and old-age benefits. Owing to this drawback, to the relatively small range of uncovered occupations, and to the fact that wages would be forced down in uncovered occupations by an exodus to those occupations, it seems unlikely that the pay-roll taxes lead many workers voluntarily to leave taxed positions for jobs elsewhere. Finally, any attempt to shift the tax forward through higher wages will meet the pressure of substitution of other productive agents for labor. In view of these considerations, it seems clear that the burden of the pay-roll tax on employees' wages falls primarily on the workers themselves.

2. The incidence of the portion of the tax imposed on the employer is also likely ultimately to be borne largely by employees, though it differs in that it is much more likely to lead to unemployment of covered workers under certain circumstances. The effect of the tax on the employer is to increase the effective wage cost. If the firm was previously in a maximum profit adjustment, after the tax it will prove advantageous to curtail employment, both restricting output and substituting other productive resources for labor. Insofar as output is restricted, part of the tax burden may be shifted forward to consumers in higher prices, especially where demand for the product is relatively inelastic. But the tax raises only the cost of labor to employers, and in view of the possibilities of substituting other productive services for labor and the lack of alternative untaxed employment opportunities for labor, except in cases of highly inelastic product demand the principal adjustments are bound to be in lower wages or unemployment, or to some extent in both. Insofar as wages are flexible, they will move down so as to permit the same total number of workers to be employed after the tax as before. Actually, however, wages are fixed by trade-unions in many cases and by minimum wage laws in others, especially over relatively short periods, and in such cases imposition of the tax is likely to lead directly to unemployment. How great the unemployment will be in any particular occupation depends

on many factors, including especially the substitutability of other agents for labor and the strength of unions involved in their attempts to prevent such substitutions. If labor markets are free and the supply of labor as a whole is relatively inelastic, the adjustment will be mainly in wage rates and the incidence of the tax will be primarily on the workers through lower wages.

Evaluation

Pay-roll taxes are frequently defended as benefit taxes. If it were regarded as desirable that workers themselves should bear the major costs of unemployment and old-age insurance, and if it were reasonably certain that wage rates were flexible enough to permit wage adjustments rather than persisting unemployment, then pay-roll taxes could be considered an acceptable means of financing social security. But neither of these "ifs" is generally accepted.

First, the regressive nature of pay-roll taxes is evident in the fact that the burden is borne in large part by the wage worker and not by the higher income groups. Those who are particularly concerned to modify inequalities in income distribution and to distribute tax burdens according to ability to pay therefore object strongly to the pay-roll tax, demanding instead at least partial contributions from other more wealthy groups in society. If proportional pay-roll taxes are to be justified it must be on the benefit principle—on this basis a case can be made for them.

Second, wage rates are far from flexible, especially where there are strong union controls or minimum wage laws. That part of the tax that is deducted from the employee's wage is likely to be accepted fairly readily by workers; the quoted wage remains unchanged. Employees, however, are likely to resist strenuously the deduction of the employers' share through lower wages. Where such resistance occurs the likelihood is that the adjustment will be through decreased employment instead of lower wages. Pay-roll taxes on the employer thus often cause directly the very unemployment that they are intended to insure against. If it is desired that the worker pay the bill, there is a strong case for putting the tax directly on him by deducting taxes directly from his wages. If it is desired that those with high salaries and incomes from property contribute, then some other kind of tax is needed. Putting the tax on the employer in the present fashion, even though we want him to pay it, has little to recommend it, since in fact the tax will almost certainly be shifted to the workers, either through lower wages or through unemployment.

Pay-roll taxes are sometimes supported as a good revenue source, quite apart from the social security program. A flat pay-roll tax of 10 per cent has received widespread support as an effective wartime revenue measure. Advocates point especially to the large sums that can be raised quickly by such a levy, collected at the source through deduction from wages. That heavy pay-roll taxes would be an effective anti-inflation measure cannot be denied, but like the sales tax they achieve this result by imposing a highly regressive burden on the lower income groups. In

certain respects they are even more inequitable than sales taxation. Pay-roll taxes fall on wage earners; income from property, from professional services, from entrepreneurial profits, and from farming goes tax-free. Perhaps heavy regressive taxes are necessary in wartime, but if so the pay-roll tax is certainly inferior to a flat-rate gross income tax falling on all groups. Opponents of all such measures argue, however, that net income taxes with exemptions, however small, for the lowest income groups could obtain needed revenue and checkmate inflation equally well and far more equitably than any flat-rate gross levy.

CHAPTER 49

Taxation (Continued)

Personal Income Taxes

PERSONAL income taxes are levied by the federal government and by about two thirds of the state governments. They now constitute the largest single governmental revenue source, being closely approached only by total corporation taxes. In the fiscal year 1938, before the recent sharp war-period tax increases, total personal income tax collections were slightly over \$1,500,000,000, of which over five sixths went to the federal government. Personal income tax collections then accounted for about 10 per cent of total tax revenues, and for about 20 per cent of federal tax collections. Since then, state income levies have changed comparatively little; but with the involvement of the United States in the war and the resulting tremendous financial needs, the federal levy has been increased drastically, both through raising rates and through lowering exemptions. In addition, collections have been increased by the rise in national income; since tax liability is stated in terms of percentages of net income received, the higher that net incomes go, the larger are tax liabilities. Substantial rate increases were enacted in 1940 and 1941, the income tax yield being virtually doubled by the tax bill of September 1941. Yet it was apparent even before the law had taken effect that much larger collections were necessary to check inflation in view of the armament program under way. Following the President's budget message for fiscal 1943 the Treasury proposed tax increases of over \$9,000,000,000, of which over one third was to come from the personal income tax, raising federal collections from that tax alone to over \$8,000,000,000 annually. This sum greatly exceeded total federal tax collections in 1938, but it was apparent, to economists if not to Congressmen, that great further increases were necessary if the inflationary pressure was to be held in check.

Under such circumstances, any detailed description of the personal income tax is certain to be out of date almost before it is written, especially since exemptions and methods of calculating net income are often changed at the same time as rates. Under the schedules in effect early in 1942, rates ran from about 10 per cent on the first dollar of taxable net income to 81 per cent on all taxable income in excess of \$5,000,000, the rates becoming steeply progressive for average families only in the income brackets above \$10,000, a category embracing less than 2 per cent of the population. Most American families have always been com-

pletely exempt from the tax. However, the bill passed in October 1942 substantially lowered the level at which real progression begins, with marginal rates of about 25 per cent on \$5,000 family incomes, and vastly increased the number of income taxpayers by lowering personal exemptions to \$624 for a special 5 per cent "Victory Tax."

By law, a tax return must be filed by every person receiving gross income in excess of personal exemptions. These exemptions were, in early 1943, \$500 per year for a single person, \$1,200 for a married couple, and \$350 for each dependent other than husband or wife. Above these exemptions, additional deductions from gross income are allowed for expenses necessary to earning one's income (such as traveling expenses for a traveling salesman), for gifts to charity and other such causes, for most other taxes paid, for uninsured losses through accidents, thefts and natural catastrophes, and for certain other costs and expenditures. In addition, the law has long contained a peculiar provision which permits deduction of 10 per cent of all "earned" income in computing taxable net income, "earned income" being very loosely defined to include at least all income up to \$3,000 per annum.¹ After these exemptions and deductions, net taxable income remains.

An additional levy effective on 1943 incomes was a special "Victory Tax"—in effect, a proportional 5 per cent net income tax on all income over an annual exemption of \$624 per taxpayer. This tax of course greatly reduced the minimum income taxpaying level, and the majority of all earners are subject to at least some tax under the levy, which is collected at the source on most wages and salaries. A portion of the Victory Tax will be refunded to taxpayers after the war—25 per cent for single persons and 40 per cent for married persons, plus 2 per cent for each dependent, except that the maximum refunds are \$500, \$1,000, and \$100 on each year's tax for the classes mentioned. If taxpayers currently buy government securities or pay life insurance or debts, they may count such payments as current deductions from Victory Tax liability up to the extent of postwar credits, in which case they of course are not entitled to the postwar refunds. This scheme of postwar refunds (or "forced saving") represented an adoption of a principle that had earlier been accepted in several other warring countries.

The regular federal income tax is divided into two parts: a "normal" tax, which applies at a flat rate of 6 per cent on 1942 taxable net income, and a "surtax," which applies at progressive rates at different levels of income. For 1942, for example, a four-person family having a gross income of \$2,500 was entitled to \$1,900 in personal exemptions and might have had about \$200 of allowable deductions, leaving as taxable \$400. On this sum they paid in 1943 the normal tax of 6 per cent on \$360 (deducting 10 per cent of taxable income under the "earned income"

¹ The original purpose of this provision was to give special consideration to earned as against unearned income. In practice, however, the primary result has been a general reduction in tax liabilities.

credit), and the minimum surtax rate of 13 per cent on the full \$400 (since the "earned income" deduction is permitted only for the normal tax). Thus, their tax bill was \$21.60 normal tax plus \$52 surtax (in addition to perhaps \$55 of net Victory Tax liability). At the other extreme, a family with a taxable income of \$1,000,000 would have been subject to the peak rate of 88 per cent (6 per cent normal tax plus 82 per cent surtax), in addition to the Victory Tax. This does not mean that the 88 per cent rate would have applied to the entire taxable income, but rather only to that portion in excess of \$200,000. On the portion from \$150,000 to \$200,000 the rate would have been 87 per cent; on the portion between \$100,000 and \$150,000, 85 per cent; and so on down to the minimum normal plus surtax rate of 19 per cent on the first \$2,000 of taxable net income. Thus the effective average rate is always less than the peak marginal rates paid. A provision of the 1942 bill also provided that in no case could taxes exceed 90 per cent of the taxpayer's income, a provision relevant of course to very few persons.

These 1942-43 rates and exemptions meant that the average family paid the Victory Tax but did not pay any substantial federal income tax unless its income was well over \$2,000; two years earlier the minimum level had been around \$3,000. Even under the drastic increases in late 1942, on a \$2,500 income the average family paid only about \$100 (including Victory Tax less postwar credit), or about 4 per cent of total income. On a \$4,000 income the tax was near \$400, or about 10 per cent of income, and on a \$10,000 income it was around \$1,900, or still under 20 per cent of income. However, on very high incomes effective rates were stringent—for example, about 80 per cent, or \$1,600,000, on a \$2,000,000 income, assuming all income to be derived from fully taxable sources (no tax-exempt securities or capital gains) and no capital losses or division of family income to minimize the tax. Actually, however, it is highly profitable for wealthy families to take advantage of all available legal loopholes, such as tax exempt securities, so that these figures substantially overestimate tax payments in the majority of cases.

Under the stress of vast war revenue needs, much pressure has been exerted for imposing almost confiscatory rates on very large incomes; President Roosevelt proposed to Congress in early 1942 taxes making it impossible for anyone to have over \$25,000 annual income after income taxes. However, even confiscatory taxation of such high incomes would absorb only a relatively small portion of total spendable income. To be an effective anti-inflation levy the tax must bear much more heavily on what are usually thought of as the "middle income groups"—families receiving from \$2,000 to \$15,000 per year—which still escape with relatively light burdens. This result could be obtained by raising rates on these groups or by lowering present exemptions. The latter method, while it draws new citizens into the income tax paying group, has the primary result of raising the effective rate on those groups already subject to tax, by lifting their incomes into the higher rate brackets.

Those favoring much heavier income taxes on the "middle income

groups" point to the still relatively light total tax burden on these groups, as indicated in the preceding chapter, and to the heavy regressive consumption and pay-roll taxes being imposed on the low income groups. While the United States federal income tax until recently surpassed the British rates on very high incomes, it has long been far lighter at the middle and low income levels. Even before the present war, England had a base rate of 27½ per cent on all taxable income above smaller exemptions than ours, and since then English rates have been heavily increased so that even following our 1942 tax bill they were heavier than ours throughout, except for a small income range where the American "Victory Tax" applies below the British personal exemptions. However large a British income, it is virtually impossible to have over the equivalent of about \$25,000 in American dollars left after payment of taxes, and a four-person family receiving as little as \$2,000 in 1942 paid over \$300 income tax.

The manner in which income tax liability is computed has long provided certain major avenues of avoidance for federal income taxpayers. At least five of these loopholes are sufficiently important to merit special mention, even in a brief description of the levy:

1. Interest on government securities, federal, state, and local, has long been exempt from income taxation, providing a wholesale avenue of avoidance for wealthy investors receiving large sums of property income. By investing in tax exempt securities such persons have been able to avoid the highest marginal rates to which their income would otherwise have been subject. Many persons, including every president since Harding, have repeatedly sought removal of this inequitable provision, but Congress has steadfastly refused to change the law, partly because of opposition from those voters benefiting from the tax loophole, partly because of opposition from state and local governments whose securities would be made taxable, and partly because of hesitation to wipe out the "vested interest" in tax exemption of present holders of tax exempts. In 1941 the Treasury ceased to issue securities exempt from federal taxation, and administration pressure on Congress to apply the federal tax at least to new state and local securities has been increased. Yet complete removal of tax exemption remains largely in the hands of Congress, which shows little inclination to change its long-standing position on the issue.

2. By dividing family incomes among different members of the family, the highest income rate brackets may be avoided and the total tax reduced correspondingly. This is easily done by transferring ownership of some of the family property to the wife or child, who then pays on his separate income at rates applicable to it, while the husband pays on his portion at rates applicable to it. Even at the relatively moderate 1940 rates, division of a \$100,000 income into two equal parts in this manner would have reduced the surtax from about \$37,000 to \$24,000, or by about 13 per cent of total income. With present higher rates, the evasion advantage is even further increased. This avenue of avoidance is also under

attack by Treasury officials, but Congress has repeatedly refused to remove this loophole for wealthy political supporters.

3. Where wealthy persons exercise control over corporation policies, they have been able to reduce their personal taxes by retaining corporate profits in the firm rather than paying them out as dividends. Since only income actually received is subject to the personal income tax and since corporations have been taxed more lightly than persons would be on income so retained, in the past this loophole has cost the Government millions of dollars annually. In some cases, special dummy corporations have been established solely to utilize this evasion method. At present, however, even though the loophole has not been closed corporation income tax rates have been raised so drastically as to reduce greatly the possibility of tax avoidance through this method. The problem of taxing undistributed profits will be more fully considered in connection with corporation taxes; one solution proposed is a special tax on such profits, another is their taxation, even though undistributed, direct to the stockholder.

4. "Capital gains" on assets held over eighteen months have not been subject to regular income tax rates. Instead, effective rates on such gains have been only 20 per cent for assets held between eighteen and twenty-four months, and 15 per cent for assets held over twenty-four months. Since under the law income becomes taxable only when realized and since the rates applicable to capital gains are so drastically below income tax rates for wealthy persons, by holding gains unrealized for over eighteen months wealthy persons can avoid a major portion of the tax that would otherwise apply. This loophole has been especially important for security investors during stock-market booms such as that of the late 1920's, although it is equally applicable to gains from appreciation of the value of any other asset.

Some economists have suggested that the same remedy for this loophole could be used as for the undistributed profits tax gap—taxation of capital gains direct to the stockholder as a part of income even though unrealized or undistributed. Others object strongly to this proposal on two grounds: first, that it is improper and inequitable to tax unrealized income, especially in view of the great difficulties that might be faced by taxpayers with heavy liabilities but no cash realizations; and second, that the administrative difficulties of acceptably determining tax liabilities on capital gains under such an arrangement would be insuperable. Many who favor plugging the tax loophole but oppose the tax on nonrealized income suggest raising the rates on capital gains to levels more nearly equal to current income tax rates so that the possibility of evasion would be greatly reduced even though the tax could be avoided until realization. Whatever the method used, there is little excuse for permitting this loophole for wealthy investors to remain open.

5. With a few minor exceptions, the federal income tax applies to money income only. Thus persons occupying homes that they own are placed at a substantial advantage relative to those who rent, especially since rent bulks so large in the average family budget. Many families

also obtain a substantial amount of real income in the form of self-produced food, but most such families are in the low income brackets where the tax would apply only slightly if at all. While failure to tax real income from owner-occupied homes (as is done in most other countries) does not provide a major avenue of tax avoidance for the very high income groups, it does provide an important loophole for the middle income groups and seems clearly unfair as between individuals in the same groups.

State personal income taxes are similar to the federal tax, with one major exception. The rates under all the state taxes are much lower, especially on the higher income brackets. In only seven states does the maximum rate exceed 7 per cent, and in several the effective rate on even very large incomes is only 2 per cent or 3 per cent. Until the heavy war-period federal rate increases, most state laws provided for exemptions somewhat lower than under the federal Act, but the \$624 federal Victory Tax exemption reversed this situation. Quantitatively, state income taxes are unimportant compared with the federal tax, but combined with the federal tax they spread the income tax burden somewhat further down to the middle income groups. There is, however, some offset, since the federal government allows a deduction from taxable income of all state income taxes paid.

Incidence

The incidence of the personal income tax is one of the easiest cases for determination. In general, it is safe to say that the incidence will be on the taxpayer; the tax cannot be shifted. Neither the demand for nor the supply of any good or service is likely to be affected by the tax to permit shifting. The prices an individual has to pay for the things he buys will not be any less simply because he pays an income tax.² Nor will he be able to get any higher price for productive services that he has to sell simply because he has to give up part of his income in taxes. Merely because an employee has to pay an income tax, he does not become any more valuable to his employer—there is no reason to suppose the employer will pay higher wages than before. Merely because a resource owner has to give up part of the funds that he receives is no reason why his resource will command a higher price or rental from entrepreneurs. The upper limit to the wage or rent paid is set by the agent's marginal value productivity, which is in no way increased by the tax.

Just as the personal income tax cannot be shifted by a wage earner or rent receiver, so it cannot be shifted by an individual in business for himself, for instance, a lawyer. Imposition of the income tax on the

² Although of course there may be a shift in relative prices because of the transfer of income from individuals to the government and hence to other individuals. The government may spend the funds on stenographers, whereas the individual would have spent them for an automobile.

lawyer does not make his services any more valuable to his clients—there is no increase in demand to facilitate shifting. Nor is the tax likely to decrease the supply of lawyers' services. If it were a special tax on lawyers, it might well cause an exodus to other occupations, but the personal income tax is a general tax applying to personal income from all occupations. If a lawyer is making his optimum profit before the tax, he will still be best off with his same fees after the tax. He must give up a certain per cent of his earnings, but the larger the profit before payment of the tax the larger will be the remainder after the tax. The tax is unfortunate, from his point of view, but it does not affect the most profitable fees for him to charge. For him to raise his fees would decrease his profits, just as it would have done before imposition of the tax.

Evaluation

1. **Effects on the distribution of income.** From the point of view of equity, the personal income tax is widely agreed to be the best of all levies, since it is a progressive tax on the basis of net income received, probably the best available indicator of ability to pay. Although the extent of redistribution obtained depends also on the expenditure policies followed, use of the income tax makes possible any desired steps toward equalization of incomes through fiscal policy. Not only does it conform to ability-to-pay criteria as levied, but we can be reasonably sure that the incidence will be where it is intended, a rare virtue among taxes. Of course, no particular degree of progression is inherent in the income tax; the steep rate structure now in effect is almost entirely a product of the last decade and especially of the war demands of the past few years. The federal tax has long been a "rich man's tax," since not more than the upper 10 per cent of the population was appreciably affected by it. War-time reductions in exemptions and increases in rates greatly broadened the range of income taxpayers, and the Victory Tax imposed a burden on nearly every income receiver. While effective rates have been increased greatly on the so-called "middle classes," these groups still fare very well compared with the "poor" in the total tax picture, contrasted with the situation in, say, England and Canada; and there appears still to be a strong case for increasing rates especially for the income levels, let us say, between \$2,000 and \$15,000. Many economists believe it would be desirable to place almost complete reliance on income taxes as the source of federal revenue, but since this is most unlikely, imposition of heavy income taxes on the low income groups, which are so heavily burdened by other levies, is not widely supported. None the less, exemptions might be substantially lowered to increase rates in higher brackets without greatly burdening those just above exemption levels.

The personal income tax, equitable as it is, presents to the wary taxpayer several avenues for escaping the rates apparently applicable to his income. Tax-exempt government securities, utilization of capital gains provisions, division of income for tax purposes, control over distribution of corporate earnings, and direct consumption of real income have

been indicated as perhaps the most important of these loopholes. Under the pressure of war, these gaps have been partially closed, but as long as such invitations to tax avoidance exist serious inequities as between individuals, to say nothing of tax revenue losses to the government, are certain to result.

With the present pattern of government expenditures, the highly progressive impact of the personal income tax results in an obviously altered pattern of resource use, not only in the shift of resources from civilian to war production but also from "luxury" to more commonplace civilian goods. While much of the income tax paid by the upper classes comes out of funds that would have been saved had it not been for the tax, such heavy taxes as now must be paid surely impinge somewhat on consumption habits, especially for the "middle" income groups. When taxes bite into incomes of these groups, the demand for fur coats, Buicks, and fine houses suffers; when armament workers receive the funds the demand for more commonplace goods and services rises. Since late 1941, however, when government priorities and allocations began increasingly to control the use of basic resources, production has been guided only to a limited extent by consumers' demands, and the effects of steep income tax rates have served only to reinforce policies primarily enforced by other methods.

2. **Effects on the level of national income and employment.** Personal income taxes are probably the least deflationary of any important revenue source that may be used to finance depression expenditures. In large part such taxes draw on otherwise "idle" funds, the expenditure of which then represents a net increase in total expenditures. By the same token, however, income taxes falling largely on the upper income groups have severe limitations as anti-inflation measures in boom periods, since they do not restrict consumption to anything like the full amount of the tax. Insofar as the income tax reaches far down into the low income groups, it loses some of its advantages as a depression tax and gains as a device for checking consumer spending and inflationary pressure. A special advantage of personal income taxes is that they may be made relatively flexible through the cycle—rates may be changed without severely disrupting economic organization or unjustifiably disturbing "vested interests." But although the income tax may be made far more flexible than many other taxes in executing stabilizing fiscal policies, even it is far from satisfactory in its present form. Collections have long lagged at least a year behind rate changes since payments have been made only after the end of the year. Collection of income taxes weekly or monthly at the source out of wages, salaries, and dividends has proved highly satisfactory in England, Canada and other countries; adoption of this arrangement would greatly improve flexibility in the United States. Especially when consumer expenditures must be held down currently to check inflationary pressures, as in wartime, is source collection critical on a much wider scale than merely under the Victory Tax.

Although the personal income tax does not represent a business cost

and restricts consumption only to a limited extent, some persons object to its use on the grounds that through the high rates on large incomes it deters investment and reduces the incentive to production. Any tax is deflationary, especially in depression periods; the proper claim for the income tax is that in such periods it is probably less deflationary than others, as well as being based more clearly on the "ability" principle—not that it is in itself inflationary or even neutral.

The extent to which high income tax rates reduce the incentive to work and to assume investment risks in periods of relatively full employment is much less clear. Much evidence indicates that as long as any substantial portion of increased earnings remain after taxes little reduction in productive activity results, since relative incomes apparently are at least as important in providing incentives as are absolute dollar levels. Especially in the high income groups, the absolute desire for increased real income is likely to be less important than "keeping up with (or getting ahead of) the Joneses." It is clearly true, however, that the federal income tax has provided a special inducement to avoid risky investments, for it takes a high proportion of very large profits but offers only a slight offsetting allowance if losses result. This criticism applies also to corporation income taxes, where it is likewise probably an important factor in deterring risky investments, especially in depression periods. This particular deterrent effect could be largely eliminated by permitting taxpayers to average together earnings in good and bad years (let us say, over a five-year period) and to pay the tax on average income. And aside from the question of investment incentives, if we tax investors very heavily on large gains it seems fair that fairly large credits should be granted them on losses that may be incurred.

In spite of its great virtues, the personal income tax is not a perfect tax. We have yet to solve the problem of taxing "capital gains" (that is, profits from increases in the value of securities and other property held). Net income is not always an adequate test of a person's ability to pay. Other loopholes in the law by which high income persons may escape the highest rates apparently applicable have been noted. Further, under present collection practices the lag in payments may be troublesome, for the individual as well as for the government. If you had a very large income last year and this year owe a large income tax on it, you may be in a very tight spot if your fortunes have suddenly taken a turn for the worse. Such tight spots arise in profusion when a depression suddenly strikes, as in 1930. Lastly, if we tax large incomes very heavily and make difficult the accumulation or maintenance of large fortunes, we are likely to find future revenues from income and death taxes greatly reduced. But in spite of these drawbacks, general agreement among economists remains that the personal income tax is the best of all taxes.

Death and Gift Taxes

Death taxes are of two sorts: estate taxes, which are levied on the entire estate of the deceased, regardless of its division among heirs; and

inheritance taxes, which are levied on the individual shares received by the various heirs. An estate tax is levied by the federal government, and all states but Nevada have either inheritance or estate taxes, with inheritance taxes slightly in the majority.

The federal estate tax in 1943 provided for an exemption of \$60,000, after which rates rose to 70 per cent on that portion of estates in excess of \$50,000,000. On the basis of these exemptions and rates, the yield for 1943 was expected to be around \$500,000,000.

State death taxes vary considerably in details, although in all cases the rates are far lower than the federal rates and in almost all the original exemptions are lower. Since the federal law provides that state taxes up to 80 per cent of the basic federal tax can be deducted by the taxpayer from the federal tax payable,³ most states have taken advantage of this provision by enacting state laws that take this 80 per cent for the state. The total collected by states in death taxes in 1938 was \$145,000,000. Thus the total federal and state death tax yield is large, but still among the smallest of the major taxes.

Both inheritance and estate taxes generally have progressive rate structures. In the estate tax these progressive rates apply to the whole estate before division; in the inheritance tax the progressive rates apply separately to each individual share after a stated exemption for each heir. For example, under an estate tax a \$100,000 estate would be subject to tax as a whole at progressive rates on that portion above the exemption. But if the estate were divided \$50,000 to A, \$30,000 to B, and \$20,000 to C, under an inheritance tax A would be taxable at progressive rates only on that portion of his \$50,000 above his exemption, B on that portion of his \$30,000 above his exemption, and C on that portion of his \$20,000 above his exemption. Clearly government revenues are larger from estate taxes with any given level of exemptions, but many persons favor inheritance taxes as a more direct personal application of the ability to pay principle.

The existence of substantial death tax rates creates a considerable incentive to avoid the tax by gifts made before death to heirs-to-be, especially since evasion of death taxes is virtually impossible when estates are transferred through the courts. To plug this loophole, the federal gift tax taxes all gifts above exemptions at rates just three fourths of those of the estate tax.⁴ The exemption provided for the gift tax is also lower than that of the estate tax—\$30,000—but an additional exception of \$3,000 annually for each donee is granted each giver. The tax is

³ This 80 per cent deduction cannot be made on specified portions of the increased federal tax recently enacted.

⁴ There is some ground for having gift rates lower than estate rates. Under the gift tax the government gets the money sooner, so that at least an interest allowance is due the taxpayer on the period for which the money has been turned over to the government earlier than under the estate tax. However, the present differential rates make it possible to avoid payment of a considerable portion of the estate tax through a judicious program of gifts before death.

collected annually, but the rates are cumulative on taxable gifts made over the entire life of the taxpayer. In 1943 the federal government expected only \$40,000,000 from the gift tax. Its importance is mainly in preventing evasion of the estate tax. Very few states have gift taxes, and the total yield of such levies is negligible.

Incidence

The incidence of death taxes may be either on the owner of the estate or on his heirs. Ordinarily the burden cannot be shifted to anyone else, for there is no transaction by which the shifting can take place. Let us note the estate tax first. In many cases, wealthy individuals take out life insurance for the express purpose of paying estate taxes. In this case, the incidence of the tax is on the owner of the estate in the form of the insurance premiums paid—at his death the insurance company will pay the taxes and the entire estate passes on to the heirs. On the other hand, if the deceased has not provided insurance for death taxes, the incidence of the tax is on the heirs, since the estate they receive is decreased by the amount of the tax. A will may be drawn so the burden is spread over all heirs, or definite sums may be bequeathed to some heirs with the residual, after paying taxes, to go to others. Under the latter circumstances the incidence of the tax is on the heirs who receive the residual portion of the estate.

Under the inheritance tax, the incidence is likewise on the heirs unless the deceased has provided insurance for taxes, which is somewhat less common than for estate taxes. The tax rate applicable to each individual heir is determined by his particular inheritance. The incidence of the gift tax, which is very similar to the inheritance tax, is generally on the recipient, unless the donor makes a gift and arranges to pay the tax himself (possibly by making the original gift larger by the amount of the tax).

It is important to remember that estate and gift taxes are exclusively "rich man's" taxes with present high exemptions. Only a very small percentage of the population ever accumulates an estate of \$60,000 to pass on at death; and the vast majority of estates completely escape death taxes.

Evaluation

Death and gift taxes are good taxes from the point of view of ability to pay. Though less directly tied to income than personal income taxes, their incidence is directly on either wealthy donors or recipients getting large sums, and shifting is virtually impossible. Both "liberals," whose interest is only progression, and "radicals," who want to eliminate inequality completely, can support such levies wholeheartedly. The plainest effect of present rates is the reduction in the size of very large family fortunes passed from generation to generation. This reduction has been rapid under recent high death tax rates, and an accelerated rate

may be expected if existing death and income levies are retained. If it is desired to redistribute income and ownership of wealth, this is an effective and equitable way of doing it in a manner that generally seems less objectionable to taxpayers than equally heavy income levies that take the funds during the lifetime of the decedent.

By decreasing the large incomes that have resulted from great accumulations of wealth, death duties also decrease savings and society's rate of capital accumulation. Such high income and death tax rates as now exist in this country act faster than is often recognized in breaking up large fortunes and incomes based on them. Some persons feel that by removing the possibility of building up and passing on huge fortunes in the future, heavy death duties will remove a major incentive to business enterprise and efficiency. Others argue that in large part it is relative position, rather than absolute wealth or income, that determines prestige and motivates the desire for large incomes and fortunes. So long as death taxes are less than 100 per cent, there remains the incentive to accumulate and pass on as much as possible above the tax burden. Surely the depressive effect of death duties is small compared with that of other levies.

It is not easy to say whether estate or inheritance taxes are preferable. If the incidence is on the owner of the estate during his lifetime, the estate tax conforms closely to ability to pay, since the rates are graded according to his entire estate. But if the incidence is on the heirs, the inheritance tax is clearly more equitable, since the rates on each share are graded according to the size of that share. The estate tax raises more income (since it applies to the undivided estates and hence higher rates are applicable) and is easier to administer, but on the equity score inheritance taxes are often considered preferable. In any case, the vast majority of estates and inheritances escape completely or pay very little in death duties.

Death taxes are obviously of little significance for fiscal stabilization purposes because their yield is so slow and uncertain, and while this yield is substantial it is small compared with that of other tax sources. For their justification death duties must rest primarily on the desire for redistribution of wealth and income in favor of the poor. On that score they rate very high.

Corporation Income Taxes

Income taxes on corporations are levied by the federal government and by almost three fourths of the state governments, such corporate levies occupying a position roughly correlative to personal income taxes. In 1938 total corporate income tax revenue was about \$1,750,000,000, of which about \$1,500,000,000 went to the federal government. Like the personal income tax, corporation income taxes have been increased repeatedly since 1938, yielding by 1943 well over five times as much as then. The basis of this huge increase was drastic increases in tax rates and soaring corporate net incomes under the stimulus of the war program.

At present corporate and personal income taxes account for almost one third each of total federal tax revenues, the remaining third being made up primarily by consumption and pay-roll levies.

The federal income tax on corporations is separated into two parts, a "regular" income tax with normal and surtax rates and an "excess profits" tax. Under the provisions in effect during 1942, all net income of corporations making over \$25,000 a year was subject to a normal tax of 24 per cent, somewhat lower rates being provided for smaller firms. In addition, a surtax of 10 per cent on the first \$50,000 and 16 per cent on all additional income was imposed. Beyond this regular income tax, the corporate excess profits tax applied at a flat rate of 90 per cent on all excess profits, except that total federal taxes on corporation income could not exceed 80 per cent of surtax net income. Two special provisions were introduced by the Revenue Act of 1942; one providing for a 10 per cent postwar refund of the excess profits tax paid, and the other providing that corporations might average out gains and losses over a period of three years in computing tax liability for any year. Moreover, the option given corporations in computing excess profits permitted many firms to escape the highest excess profit rates. Excess profits could be measured as any earnings in excess of 95 per cent of average earnings over the period 1936 through 1939, or as any earnings in excess of 8, 7, 6, or 5 per cent of invested capital for small to very large firms, respectively. In effect, this option meant that during the rearmament boom of the early 1940's many booming industries, by taking advantage of the average earnings option, made huge returns on invested capital without becoming subject to peak tax rates since they had also shown high earnings in the preceding four years. This was especially true of such heavy industries as steel and metal products, which had early felt the effect of European war demands.

Like personal income taxes, levies on corporate income have been rapidly pushed upward under the impetus of the need for war revenue, through increases in both surtax and excess profit rates. Such drastic corporate taxes are advocated especially in wartime because many of the largest profits are shown by producers of war materials. It seems equitable, therefore, to many persons that such war profits should be recaptured to prevent the enrichment of "war millionaires." Furthermore, drastic rates on excess profits appear to be a necessary basis for any concessions of wage-earning groups who are asked to refrain from demanding higher wage rates that would increase inflationary pressures.

The personal income tax is a levy on "personal net profits," but while the corporate income tax is likewise applicable to net profits as defined by business accounting methods, it is not a tax solely on "economic profits," a fact significant for the analysis of tax incidence. Taxable corporate net income includes all profits made by the corporation, before dividend payments. Interest payments on bonds are a deductible cost; dividends on stocks are not. In preceding chapters it has been emphasized that interest on proprietors' investments is as real an economic

cost as interest on borrowed funds in determining "pure" or "economic" profits. Thus when stockholders buy shares they receive their interest in the form of dividends, part of which must therefore be charged as a cost of doing business. When a corporation earns a net taxable income but this income is not large enough to pay stockholders dividends equal to the going rate of interest on their investment, the income tax must be paid even though the corporation is making an "economic" loss. The general effect of this method of tax calculation in encouraging bond rather than stock financing is obvious.

The rates of state corporate income levies are all much lower than the federal rates, though together with federal taxes they sometimes approach a 100 per cent marginal rate on very large incomes. However, the federal government permits deduction of state taxes paid in computing federal liability, so effective rates can never exceed 100 per cent. Considerable diversity exists among state laws, but the quantitative importance of such taxes is not sufficient to merit a detailed discussion of their differences. Many of them are franchise taxes, using net income only as a measure of the value of the corporate franchise to do business during the year. In general, however, their effects are similar to those of the federal tax, but on a greatly reduced scale.

Incidence

It is convenient to consider the incidence of corporate income taxes separately in the short and long runs.

1. In the short run it is reasonably certain that corporation income taxes will not be shifted, no matter whether the corporation holds a strong monopoly position or operates under purely competitive conditions. If a corporation is operating at its maximum profit output and price before imposition of the tax, this optimum output and price will be unchanged after the tax. Neither demand (marginal revenue) nor marginal costs are changed by the tax, and the same output and price would still give the maximum profit before payment of tax. Unfortunately for the corporation stockholders and executives, part of those profits now have to be paid out as taxes before funds are available for dividends; but the amount left for the stockholders will be more if profits before taxes are at a maximum than if profits before taxes are at any lower figure. The corporation is in the same position as the lawyer considered in connection with the personal income tax. The tax does not make it profitable for the enterprise to reduce output and charge a higher price nor does it enable the enterprise to pay lower prices for its productive services. In the short run, the corporate net income tax is borne by the stockholders, through lower dividends and/or through reduced stock values, and quite possibly by corporate executives whose bonuses are based on net earnings.

2. In the long run part of the tax may be shifted, but most of it is likely to remain on the stockholders and corporation executives. As in the

short run, there is no reason to suppose that demand for the corporation's products will increase as a result of the tax; therefore any forward shifting must come through a decrease in supply. There is no reason to suppose a change in the supply of productive services offered as a result of the tax; therefore any backward shifting that is to come must come through a decreased demand by corporations for productive services. A decreased supply of finished products or a decreased demand for productive services, making possible respectively forward or backward shifting, might come either (a) through a shift of investment away from corporations to other business forms or (b) through an outright withdrawal of investment from business and its use for consumption. If the tax is to be shifted one of these must occur.

a. What likelihood is there that investment will shift away from the corporate form into other forms of business organization, thus eventually raising prices of corporate products and lowering prices paid for corporate-used productive services? Such a shift might occur if the net yield on corporation stocks were forced below the going interest rate by the special corporate tax. But probably there is comparatively little likelihood of a shift away from corporations to other forms of business organization even though corporate income is sharply taxed, for two reasons. First, other forms of business organization are also taxed. Second, the corporation form is almost essential in financing and operating large-scale enterprises. The fact that in most cases such a shift away from corporations has been either unnecessary or impractical is evidenced by the continued and ever-increasing investment in corporate enterprise in spite of the high yields of the corporation income tax.

b. What likelihood is there that investment will simply be withdrawn and consumed because of the corporation income tax, thus eventually raising prices of corporate products and lowering prices paid for corporate-used productive services? This conclusion might follow from the above conclusion that there is not much opportunity for investment to move into other forms of business organization to escape the tax. Especially might this be true if taxes forced down the rate of return to stockholders considerably below some going rate of return to which they have become accustomed. But how many people would in that event simply withdraw their corporate investments and use the funds for consumption purposes cannot be told. Not many, from all available evidence. The fact is that we have continued to save large proportions of the national income except in very bad years, and a large amount of this saving has gone into corporate stocks. How much, if any, lower this rate of savings and investment has been than it would otherwise have been because of the burden of the corporate income tax on the most practical form of investment, no one knows.

In general, therefore, very little of the corporate income tax is likely to be shifted either forward or backward, even in the long run. Since all investment opportunities are taxed and since a shift from corporations, even if they are higher taxed, is usually impractical, shifting, which

depends on movement of investment away from the taxed area and a reduction in corporate output and employment, is unlikely to result.

While the tax is generally not shifted to productive agents or consumers, it does encourage the use of bond rather than stock financing. Since interest on bonds constitutes a deductible cost while dividends do not, by shifting from stock to bond financing corporations can minimize their income tax liability. Some evidence exists that such a shift has been expedited by the corporate income tax in its present form, but again there is no accurate way of telling the quantitative importance of the tax factor. There are so many other reasons why people prefer stocks to bonds, or vice versa, that it is impossible to isolate the importance of the tax factor. We can say only that, other things equal, the corporate income tax leads to a shift in corporate financing from stocks to bonds.

Evaluation

1. **Effects on the distribution of income.** Taxes on corporate income conform fairly well to the criterion of ability to pay. The burden is largely on corporate security holders, and most such securities are held by the upper and middle income groups. However, the corporate income tax falls considerably short of the personal income tax on this score. The corporate tax, to the extent it is an ability-to-pay tax, must assume that income from securities is a satisfactory measure of taxpaying ability. This may or may not be the case. Mr. A, who has an income of only \$2,000 annually, may own a few shares of General Motors. He will lose perhaps as much as 75 per cent of his potential income from dividends because of corporate income taxes. The millionaire owner of General Motors stock loses exactly the same percentage through the tax. Though in general corporate income taxes fall primarily on well-to-do people, they represent only a rough application of the ability-to-pay principle. "Ability to pay" is a concept that applies only to persons, and the personal income tax is a far more direct application of this principle than taxation of business income that combines the incomes of many persons of diverse economic positions.

The heavy excess profits tax has been widely favored as an especially equitable type of income tax which helps close the loophole in the personal income tax through retaining corporate earnings in the firm. That this tax is, however, at best an unsatisfactory makeshift in patching up personal taxes has already been indicated. The failings of the tax from the equity standpoint are greatly increased by the discriminatory method of calculating excess profits subject to tax, which permits many high income corporations to escape through choosing the average earnings option.

The benefit principle has also been invoked in support of corporate income taxes. By granting a corporation a charter the government bestows upon it certain privileges not available in other forms of business organization. It is therefore regarded as equitable that the government should recoup some of the fruits of this special privilege through taxation.

This argument appears to have considerable validity, but it ought not to be confused with a tax on the general benefits provided by government to all types of business. A special tax on corporate income can be justified on the benefit principle only insofar as the benefits are special to corporations. If the benefits bestowed are general to all businesses, the tax based on them should also be a general tax on all businesses.

2. Effects on the level of production and national income. Taxes on net income impose no burden on marginal costs and hence do not have the same direct deterrent effect on production and employment as do excises and sales and pay-roll taxes. On the other hand, the level of investment depends on profit expectations, and a tax that absorbs a large percentage of any profits above what are considered "normal" levels necessarily removes a considerable portion of the inducement to invest. Especially is this deterrent effect likely to be important in cases where risk bulks large. Under existing laws a very large proportion of any extraordinary profits are certain to be absorbed by taxes, but the compensating credits allowed in case of losses are far more restricted. Naturally, investors tend toward more conservative, low-yield investments under such tax arrangements. While this tendency may be thought generally desirable, in depression periods it is essential that "risky" investments be undertaken if full-fledged recovery is to be obtained, and a tax structure that specifically penalizes such investments is hard to defend. The pressure of present tax arrangements toward a shift from equity to bond financing is likewise felt by many persons to be undesirable because fixed obligation financing is likely to intensify any deflation that may develop. When businesses have fixed cash charges to meet they must liquidate enough assets to meet the obligations or accept bankruptcy, however ill-fitted the period may be for such liquidation. On the other hand, a business financed primarily by stocks is free from the pressure to meet bond interest payments during deflationary periods. Given equity financing, when earnings are made they can be paid out, but if no dividends are paid there is no resulting bankruptcy.

In an inflationary period such as recent years the corporate income and excess profits taxes may have important fiscal-political functions in addition to the actual funds that they transfer out of the income stream to the government. Without very heavy profits taxes it has proved politically impossible to obtain effective restrictions on other inflationary pressures, particularly personal incomes received by wage earners and farmers, the two major political pressure groups in recent years. Thus the importance of steep excess profits taxes in checking inflation may be all out of proportion to their actual fiscal magnitude if they help make possible effective controls on spendable incomes of the middle and lower income groups where the great bulk of total spendable income is received

The Taxation of Property

Property taxes have, until the present war, been far the most important single tax source in the United States, accounting up to 1941 for roughly

one third of all taxes collected. They are also the oldest major source of tax revenue. The federal government levies no property taxes, and state governments now receive only about \$200,000,000 annually from this source (though property taxes were long the most important state revenue source). But local governmental units obtain almost \$5,000,000,000 annually—virtually their entire income—from property taxes.

Property taxes fall into two major categories: taxes on "real property"—that is, "land" and improvements on the "land"; and taxes on "personal property"—that is, property other than "land" and improvements on it. A further subdivision of personal property is often made into "tangible" personalty, such as furniture, automobiles, machinery, and clothing, and "intangible" personalty such as stocks and bonds, money, bank deposits, and mortgages.⁵ Property taxes vary considerably as between various taxing units. Some governmental units tax only real property; others tax all property at identical rates; others exclude intangible property, or apply a special low rate to it. Moreover, the administration and incidence of taxes on various types of property is likely to be quite different.

Administration

How does the property tax actually "work" in practice? The general practice is for local units—towns, water districts, school districts, counties, road districts, and the like—annually or biennially to estimate the expenditures required for the coming year or two years. These estimates, when finally accepted by those in control, constitute the amount to be raised, and the major source is the property tax. Meanwhile, the property to be taxed has been assessed (anywhere from annually to once every five years) by assessors hired by the government, and the assessed value of the property placed on official "tax lists." Then the amount to be raised by property taxes is compared with the assessed property values shown by the tax list, and the tax rate necessary to raise the funds is determined. For example, if \$10,000 is to be raised and the total assessed valuation is \$1,000,000, then a tax rate of 1 per cent on assessed valuation is necessary to raise the funds. (Often this rate is quoted as 10 mills, to show that the tax would be 10 mills on each dollar of valuation. Sometimes it is quoted as \$1 per hundred of assessed valuation.) The tax is then levied and collections made by the tax collector.

Actually, the administration of this process has been shamefully inefficient. Even in the case of real property, assessments have been indefensibly bad over a large part of the country. Land and houses cannot hide. Yet in numerous cases efficient assessment surveys (often now conducted by aerial photography) have revealed large numbers of rural and urban properties that have never even been listed for taxation. Furthermore, even for property that is assessed there is a tremendous variation in the level of assessment. In almost all cases it is true that

⁵ In general, intangible personal property consists of documents showing ownership of, or claims on, wealth or income, while tangible personalty is wealth itself.

low-priced real property is overassessed relative to high-priced real property (that is, assessment is "regressive")—largely because most assessors have no idea of the value of expensive properties and hence depend on the figures suggested by owners or simply follow past assessments. In a recent Texas study it was found that assessments of real property varied from less than 10 per cent to over 110 per cent of actual sale value, with the largest percentage of the cases falling in the 20 per cent–30 per cent bracket. So long as all property within a taxing jurisdiction is undervalued by the same percentage, no injustice to individual taxpayers is done, since the tax rate is simply that much higher per dollar of assessed valuation in order to raise the required funds. But when some property is assessed at a larger per cent of actual value than other property in the same tax district, the more highly assessed property receives more than its share of the tax burden.

Bad as the assessment of real property is, that of personalty is far worse. Here the results have been ludicrous in many cases. The situation in Wisconsin before the reforms carried out by a newly created tax commission is indicative. One year eighteen counties listed no grain or farm products; twenty-three counties listed no household furniture above the \$200 per house exemption; one county listed not a single watch. Very recently, the Federal Bureau of Public Roads estimated that about 40 per cent of all automobiles susceptible to property taxation have never even appeared on the tax rolls, much less been assessed accurately. And in assessment of intangible personalty, the results reach the extreme of futility. There have been numerous cases where entire counties have failed to show any bank deposits on the tax lists, in spite of the fact that numerous banks were operating openly within the county. In Cook County, Illinois (Chicago), the per capita listing of money one year was 64 cents and that of credits 45 cents, in spite of the fact that bank deposits alone totaled far over a billion dollars, to say nothing of other credits and actual cash. In one famous year (1912), Kentucky obtained more revenue from the dog tax than from the tax on all stocks, bonds, money, and credits.

The assessment of personal property is faced with certain difficulties that are more serious than for real property. Real property cannot hide, but many types of personal property are relatively easy to conceal. This is especially true of intangibles, for the assessor has no right to invade your safe deposit vault nor to pry into your personal affairs in the hope of uncovered unlisted intangibles. Part of the bad assessment results may be excused on such grounds, and part of the unequal assessments may be explained by the fact that accurate assessment is tremendously difficult, especially for large industrial plants and utilities. But the major blame must fall on inefficiency and the lack of real effort to enforce the laws. In many cases local administrators make little pretense of enforcing statutes requiring taxation of all property. This is especially true in regard to intangibles, where often all pretense of enforcement has been given up. It may be impossible to assess some sorts of property in accordance with existing laws, but there can be little justification for

simply disregarding the law. If we are unable or unwilling to enforce existing laws, surely the thing to do is to repeal them or re-enact them in a workable and acceptable form. This latter step has finally been taken by many governments—intangibles have been legally exempted from taxation, or a special low rate has been applied to them in the hope that owners will voluntarily list intangibles for taxation at the lower rate.⁶

In many areas the collection of the property tax, once levied, is hardly better than assessment procedures. In most cases, local tax collectors are of the same caliber as the assessors, and the "collection" process often consists solely of publishing a notice that taxes are due and sitting in the collector's office to await the taxpayers. As a result, often a large percentage of taxes levied goes uncollected, especially in depression years. The worst major example was Chicago during the depression of the 1930's—even as late as 1938 almost \$250,000,000 of uncollected property taxes were outstanding. Partially this was due to economic conditions, but also in large part to an inefficient collection system. At the other extreme are cities like Cincinnati and Milwaukee, which collected 95 per cent of their total levies on time in the same year. To the extent that part of the levy is uncollected, or late in collection, nonpayers escape or postpone their share of the burden, and the operation of the public economy is hampered. Property tax administration is not inefficient everywhere, but the areas of marked inefficiency are so prevalent that it is necessary to recognize these inefficiencies specifically if we are properly to analyze and evaluate the property tax. No other major tax is open to anywhere near as much criticism on this score.

The prevalence of inefficient administration of property taxes has important results. Assessments tend to be regressive; assessment procedures have been especially susceptible to manipulation and graft, to "arbitrary" action of government officials and powerful private individuals; at bottom, poor administrative procedures often reduce to the inefficiency of untrained, politically appointed local assessors coupled with a public attitude that it is quite justifiable to evade these taxes "if you can get away with it." The inequities and uncertainties in assessment and collection procedures are important factors to be taken into account in evaluating the effects of various types of property taxes; tax incidence analysis must take into account administration differentials if an accurate picture is to be obtained.

Taxes on personal property

Recognizing the existence of these administrative weaknesses, the incidence of personal property taxes may be considered briefly. The property tax on household goods, clothing, and other belongings used for

⁶ This exemption or special rate on intangibles is also sometimes justified by the argument that taxation of intangibles at regular rates may result in serious double taxation. For example, if a mortgage is taxed on its full value and the mortgaged farm is also taxed at full value, the same property is being taxed twice under the property tax.

direct consumption obviously cannot be shifted, for there is no later price transaction through which the shifting could take place. If consumers anticipate the tax, they might be willing to pay less for purchases, but to the extent that the tax applies equally to all property held (purchases made) there is little reason to suppose that consumers' demands for various products would change.

The tax on personal property used in business (that is, on office equipment, machinery, and raw materials) may or may not be shifted. Insofar as this tax applies to all businesses (and almost all businesses use some personalty), there will be no place else for investment to shift to escape the tax, and the effect will be simply to reduce the net earnings and "going value" of all business enterprises. This process will be discussed more fully in the following section on "real property," where it is characterized as "capitalization." The burden of taxation on personal property, however, varies substantially between different business enterprises. To the extent that this variation is important enough to lead to shifting of resources, the tax on business personalty will lead to a shift of investment away from businesses using much taxable personalty to those using little. Actually the personal property tax, as administered, is probably not sufficiently important in the calculations of most business enterprises to lead to such shifts on any very large scale. Where it does reduce the production of any good through inducing shifts of investment, the tax will be to that extent shifted forward to the consumer or backward to owners of productive resources, depending on the relative elasticities of product demand and resource supplies and on the degrees of monopolization in the various markets concerned. The general principles of shifting are similar to those discussed in connection with excises, and they need not be restated in detail here.

Taxes on "improvements on real property"

The immediate burden of a new tax on improvements will be on the owners of such property—houses, store buildings, factory plants. There is nothing in the tax itself to increase demand for these improvements (either when rented or purchased outright), or for the goods that they produce; there is no reason to suppose that the existing quantity can be sold or rented at a higher price just because of the tax.

Given time for adjustment, however, the tax will ordinarily be shifted either forward or backward. The tax burden will discourage the production of new "improvements." Some substitution of nontaxed productive agents (especially human labor) for "improvements" used in production may occur. Owners of agents specialized to various aspects of construction will either have to take lower returns or find their services partially unemployed as construction activity declines. Insofar as substitution of nontaxed agents is possible, consumers may face little increase in prices. On the other hand, to the extent that nontaxed investment opportunities are available elsewhere, improvements such as buildings and fixed machinery will not be replaced and investment will be

shifted, reducing production and raising prices of both improvements and final products if the improvements are themselves productive resources. The general analysis of shifting is identical in principle with that discussed in connection with excises. Since ultimately the quantity of improvements will be less than if there had been no tax, the price of the improvements will be increased, thus shifting part of the tax burden forward to purchasers and renters of improvements and on to final consumers. But those owning taxed property at the time of imposition of a new tax will only gradually recoup part of the tax from buyers or renters. Since such property is ordinarily very durable, the new units appearing on the market in any given month or year constitute only a small proportion of total offerings. It takes a long time for curtailed production and depreciation of old property to result in markedly reduced offerings and hence higher prices. In the short run the tax may be borne entirely by the improvement owners. But in the long run, a differential tax on some investment opportunities is shifted as investment is transferred to more attractive fields. Thus heavy real property taxes on improvements constitute a major barrier to new construction and other improvements on "land."

There is one special case in which a tax on improvements is not shifted, even in the long run. The tax on owner-occupied homes cannot be shifted (except possibly through higher rents to roomers) because the owner-occupier has no price transaction through which to shift the burden, and such taxes account for almost \$1,000,000,000 annually. The case is analogous to taxes on personal property used in direct consumption, such as clothing and furniture.

Taxes on real property ("land")

Precise separation of "land" from "improvements" on it is virtually impossible. A piece of farm "land," for example, is ordinarily soil partly built up by natural forces over centuries and partly by a variety of direct and indirect human activities, such as use for farming purposes, application of fertilizer, irrigation, and so on. To a considerable extent the amount of different types of "land" is variable by human activities. Insofar as the amount of "land" of any type is so variable, a tax on it will have the same sort of effect as a tax on such improvements as buildings. If other, less heavily taxed investment opportunities are available, investment will gradually be shifted away from "producing" or building up taxed real estate. As this shift of investment continues, the amount of "land" will decline, its price and the price of its products will rise, and gradually the tax will be shifted, even though the process may be very slow in view of the relatively inelastic supply of most types of "land."

On the other hand, insofar as the amount of "land" of any type is completely outside the control of human activity, the tax obviously cannot be shifted and must be borne by the owner at the time of imposition. The tax reduces the amount potential buyers will pay for the "land"; it does nothing to increase demand for the services of the "land."

Since the amount of the given type "land" is by assumption fixed, investment cannot be withdrawn through shifting depreciation allowances to other fields, as may be done with improvements, and there is no alternative for the owner but to bear the burden of the tax.

Since every piece of "land" is to some extent fixed in quantity and is to some extent subject to control by human activities, the net result of a new "land" tax in any case is almost certain to be a partial shifting and a partial final imposition on the owner. In the case of farm land, where soil maintenance and building practices play an important role, a heavy differential tax would be likely to be shifted in considerable part. On urban sites and mines and quarries, where current depreciation subject to replacement or shift to other investments is very limited, shifting is much more difficult. In all cases, since the range of less heavily taxed investment opportunities is definitely limited, the effect of the tax in decreasing the supply of "land" is greatly lessened. To the extent that shifting does not occur, the tax is likely to be "capitalized" on the present owner, in a manner to be considered presently.

Severance taxes

Severance taxes are taxes imposed on the value of raw natural resources as they are obtained from their natural state. Such taxes, for example, have sometimes been imposed on timber as it is cut and on ore and oil as they are extracted from the ground. Though relatively unimportant in fiscal terms, such levies have marked effects on the utilization of resources that merit special mention.

There is no way whereby the owner of natural resources can shift a severance tax imposed on his property. The tax does not increase the demand for the timber or oil that he owns. He cannot shift the burden backward onto productive agents that he hires, or onto future buyers of the property, since all future buyers will presumably take into account the necessity of paying the tax in determining the price that they are willing to offer. The incidence of the severance tax, like that on the fixed supply portion of "land," is on the owner and cannot be shifted, even in the long run. However, while property taxes tend to speed up the use of natural resources such as forests (since the sooner the forest is cut the shorter is the period for which taxes on the trees must be paid), severance taxes have exactly the opposite effect. Since the tax is imposed only when the severance of the resource occurs, the owner is likely, rationally or irrationally, to delay as long as possible to avoid having to pay the tax.

Capitalization

If a real property or severance tax cannot be shifted, the net return on the property is correspondingly diminished. Since the price that any buyer is willing to pay for a productive agent is determined largely by the capitalized value of the future net incomes that he expects to receive from it, buyers will be willing to pay less for the land after the tax has

been levied; and the capital value of the land is thus decreased by the imposition of the tax. Under such circumstances, therefore, the tax is often said to have been "capitalized." Any future buyer of land on which a property tax has been levied will take into account the necessity of paying the unshiftable property tax and will not invest in the land unless the price is low enough to provide a net return after taxes equal to that obtainable on investment elsewhere.

The crucial fact about capitalization is that, through forcing him to pay the unshiftable tax and forcing down the capital value of his property, it puts the burden of the capitalized tax exclusively on the owner of the land at the time that the tax is levied. Any future buyer will not bear the burden of the tax, because he has anticipated the tax in the price paid for the land and hence realizes a return after payment of the tax equal to what he could have received on nontaxed or less heavily taxed investments; future owners buy "tax-free."

Examples of capitalization may be found in other fields as well. Suppose that a perpetual bond with a face value of \$1,000 yields 5 per cent and the going interest rate is also 5 per cent. Then the market value of the bond will be \$1,000. Now suppose a *special*, nonshiftable tax of 2 per cent of the face value of such bonds is imposed. This will reduce the net yield from \$50 to \$30 annually, but will not affect the yield of other investments. Therefore investors will be willing to pay only three fifths as much as before, or \$600, for the taxed bond. At a price of \$600, the net return of \$30 on investment in the bond is 5 per cent, just equal to the return obtainable elsewhere. Obviously investors will not now be willing to pay more than \$600 for the taxed security—the special nonshiftable tax is capitalized. Any buyer at \$600 in effect buys free of the burden of the special tax since he receives his 5 per cent on the investment.⁷

Although actually capitalization does not always work out so precisely, it is of great importance in connection with the real property tax. It is of special significance in considering proposals to decrease the tax on real property. If present buyers have bought at lower figures because of the tax (thus capitalizing the burden on earlier owners), to lower the property tax now would be to give a subsidy to present owners through increasing the capital value of the land. A good deal of the present argument that "land" is too heavily taxed is therefore invalidated by the fact that much of the "land" tax has long since been capitalized—heavy property taxes have long been prevalent and most buyers have bought expecting the taxes to continue. This is not to say that the property tax has in all cases been entirely capitalized, but rather that much capitalization has taken place and that this fact ought to be given proper weight in considering the common argument that "land is too

⁷ This analysis assumes that the interest rate remains unchanged, a justifiable assumption for the case of a special tax that reaches a fairly small percentage of investment opportunities.

heavily taxed." The fact of capitalization is a strong argument against drastic reductions in long-established property tax rates.⁸

To summarize: If a long continuing nonshiftable tax is levied on one group of investment opportunities ("land," bonds, or whatnot) but does not fall on other investment opportunities, it is likely to be capitalized through forcing down the capital value of the specially taxed investment opportunity until the net yield is the same as in less-taxed fields. Capitalization will take place only when (1) the taxed investment opportunity is a continuing (durable) source of income, (2) the tax is expected to continue over a considerable period, (3) the tax is nonshiftable, and (4) some investment opportunities are free from the tax, the levy falling differentially on different investments. Actually, the process of capitalization is rough and inexact, especially since buyers are never sure that present tax rates will continue.

Evaluation of property taxes

1. **Effects on the use of productive resources.** Insofar as property taxes impose differential burdens on some types of investment opportunities (for example, on farming that uses a large amount of relatively heavily taxed land), they impede the allocation of resources to those industries and help induce their transfer elsewhere. In particular, heavy taxation has often deterred investment in new housing and business construction. However, since the possibility of shifting investment to relatively untaxed opportunities is greatly restricted by the scope of property taxes and of other taxes where property levies are light, such shifts of resources may in fact be relatively small. On the other hand, differential rates of property taxes in different areas and on different types of property, whether through differences in tax rates or in assessment practices, are widespread, and to the extent that these differentials are important they lead to a different allocation of resources than would otherwise have occurred. Such differentials, for example, are often important in determining the spread of housing construction in urban and suburban areas and in the location of factories.

⁸ Another important practical case of capitalization is that of tax-exempt government securities. Since the income from these securities is entirely or partially tax-exempt, their net yield to the investor per \$100 of face value is greater than that of taxable securities. This fact is capitalized in the higher prices of such government securities. Buyers of the securities, to the extent that this capitalization has taken place, obtain no higher yield than on other comparable investments, aside from progressive income tax complications. If, suddenly, after the long practice of tax-exemption, the government were to declare income from such securities taxable, this fact would be capitalized in the form of lower prices obtainable for government securities. Thus the burden would be placed permanently on present holders. This, of course, is no argument against the taxation of income from future issues of government securities, nor is it an argument against making income from present bonds subject to any income tax *increases* that occur in the future; but it is cited by many persons as a major inequity that would result from removal of outstanding exemptions.

Not only do differential taxes on property discourage new investment in improving property, but they increase the incentive to exploit such property faster than would otherwise be the case. This is particularly true with forest land, quarries, oil pools, and other such property where the tax assessment value declines apace with the extraction of the natural resource; it is much less true with farm land and urban property where assessments are much less variable depending on the rate of use. Severance taxes yield exactly the opposite result, slowing down the rate at which natural resources are used, since the tax need not be paid until the severance takes place. Whether property or severance taxes are preferable for "exhaustible" natural resources depends on how the needs for conserving resources for the future are weighed against present use. If it is desirable to leave the rate of use to consumers' choices as reflected through the price system (especially by the interest rate), neither special property taxes nor severance taxes are called for.

Representing as they do a heavy and lagging fixed money cost in depression periods, real property taxes reinforce the cumulative downward spiral of employment, production, and income in deflation periods, especially in the agricultural sector where such taxes constitute a major cost item. If assessments were revised promptly following value changes this pressure would be lessened, but under present practices assessments remain high long after property values and product prices have fallen. A related objection to the property tax is voiced by economists who feel that the level of construction activity is a primary determinant of the level of national income and employment and argue therefore for reduction or complete removal of present taxes on improvements in order to avoid restraining new building. With the general belief in a shortage of investment opportunities during the 1930's considerable emphasis was placed on this type of tax reform; with the shift of emphasis to inflation prevention in recent years and the impossibility of constructing any except the most essential structures the case for such tax reductions has largely vanished until after the war. At present the result of such a policy would be largely a windfall to present property owners in the form of lower taxes than expected at the time of purchase.

2. Effects on the distribution of income. The general property tax is sometimes defended on the ground that the extent of a person's property is a good indication of his ability to pay taxes. Even if all property were reached by the tax, this guide to ability would be a very weak one, since the major portion of national income is received in wages and salaries that have no necessary connection with property holdings. But in fact the property tax is largely only a real property tax; the vast amount of wealth held in the form of tangible and intangible personalty, such as securities and bank deposits, completely escapes taxation in many areas and partially escapes in most others. Thus in practice, even if property ownership were a good indication of ability to pay, the current sort of tax would be highly unsatisfactory. Lastly, since the long-standing property tax is to a considerable extent capitalized on

much property, often there is no incidence in a meaningful sense on present property owners. In the face of all these considerations, it is difficult to accept the property tax as a satisfactory levy if ability to pay is a major criterion. This weakness of the tax is increased by the fact that it is at best proportional on property reached and is in fact generally regressive since assessments usually more nearly approach full value on low value property.

The property tax can be justified somewhat more strongly on the "benefit" principle. Property owners receive many direct services from local governments and it seems proper that the users should pay for these special benefits. This is especially true in connection with the real property tax—owners of real property receive important special benefits, such as police and fire protection, and the larger the property holdings, the greater these benefits received.

Yet to the extent that property taxes are not ultimately paid by present property owners the justification of property taxes on the benefit score carries less weight, although if taxes are shifted to renters the benefit principle still applies, since renters obtain many of the same benefits of local government that accrue to property owners. A still broader "benefit" justification for real property taxation is advanced by those who feel that all real property should belong to the state. In the absence of state ownership, taxes are justified as a sort of rental for the use of the property.

Since the property tax and its arbitrary, inefficient administration are open to substantial criticism and offer few advantages over direct personal "ability" taxes, most economists feel that it is desirable to avoid further increases in property levies. On the other hand, it appears undesirable to reduce drastically tax rates on "land," in the narrow sense, since to do so would largely result in subsidies to present owners through capitalization. There will probably be much to be said for reducing taxes on buildings and other real property improvements after the war if it becomes desirable to stimulate construction in general and especially better low-cost housing. Many people feel that we would also do well to replace the personal property tax with other better taxes insofar as possible at the earliest convenient time. Even at their best, personal property taxes on individual and business properties are hard to defend on ability or benefit criteria, and when the difficulties of equitable and efficient administration are considered it is difficult to argue for their continuance when better taxes are available.

Special Assessments

Closely related to taxes on real property are special assessments. When special improvements (such as sewers or paving in a new subdivision) intended primarily to benefit restricted areas are installed by a local government, they are often financed by "special assessments" on property in the benefited area. The burden of payment is placed on

those property owners specially benefited, since the improvement is primarily for the benefit of that particular group. The cost of the improvement may be divided among property owners in any way that seems best to measure relative benefits received—for example, according to the value of property owned or according to street frontage, in the case of paving. Although the process of equitably determining benefits and allocating special assessment charges is not easy, such assessments do represent that element of the general property tax most directly justifiable on the benefit principle. Unfortunately, in only a few cases can the special assessment method be used satisfactorily since the benefits from most government services are very widespread and impossible to allocate accurately.

The main characteristic of special assessments is that the funds raised are used directly to benefit the property taxed. Since the taxed property is thus "improved" (perhaps by having new paving in front) the demand for the property and its rental and sale value will presumably be increased thereby. While the immediate incidence of the tax may be on the present owners, later the charge is likely to be paid by renters or buyers of the taxed property who in turn gain direct benefits from this higher price or rent paid.

Ideally, therefore, financing improvements by special assessments is very similar to direct sale of government services to consumers (citizens), with citizens' sovereignty the ideal to be achieved. Actually, the process works out very imperfectly. Often special improvements are badly planned and do not result in property value and rental increases sufficient to merit the expenditures made. And as with regular property taxes, assessments of benefits received are likely to be unfair and arbitrary as between different properties. In order to be completely justifiable on the benefit and consumers' choice principles, it is necessary that special assessments be administered efficiently and nondiscriminately as well as that they be in accordance with the wishes of owners of the benefited properties. In some cases, of course, benefits conferred by government improvements greatly exceed the costs of the improvements. Whether in such a case the government should levy an assessment only sufficient to cover costs or one high enough to recoup all the special benefits of the improvement is primarily a question of ethical principles. If the benefits from the improvement do not prove sufficient to cover its costs, a like problem arises. In general, the tendency has been to levy special assessments on a cost-of-the-improvement basis. In view of these difficulties and since a substantial element of compulsion on the individual property owner is often involved, special assessments are only a very imperfect substitute for prices in allocating the benefits and costs of improvements to real property. In respect to the compulsion involved, such levies are similar to pay-roll taxes to finance social security "benefits" that the individual worker may not want to "buy." None the less, special assessments are generally considered the fairest way of financing government-provided special improvements on real property.

Conclusion

It is useful to pull together the major points of the preceding long discussion of particular taxes in considering how a "model" federal-state-local tax system might look if formulated on the basis of the criteria suggested at the beginning of Chapter 46. Such a "model" tax system might have four major revenue sources. *First, and of major quantitative importance*, there would be a progressive personal income tax plus progressive death and gift taxes, all of which conform closely to commonly held criteria of ability to pay and which together meet most of the tests of "good" taxes. The income tax should be collected insofar as possible "at the source," and might well also include a prorating of undistributed corporate profits. Rates on such personal levies would be high enough to provide the bulk of all federal and state governmental revenues. *Second*, there would be a proportional net income tax on all businesses, justified largely on the basis of benefits received and providing much less revenue than personal levies. In view of the benefit basis for the corporation tax, it might be desirable to impose a minimum tax (possibly based on capital investment or gross sales) that would be collected even though a business made no profits in a given year. The rates would be proportional, not progressive, since the business tax would be a benefit rather than an ability levy.⁹ *Third*, taxes on *real* property, especially on "land," would be continued at roughly present levels, not so much because the real property tax is a "good" tax (although it has some justification on the benefit score) but because it is a major revenue source for local governments and one to which economic life has become well adjusted so that its removal would mean essentially a subsidy to present real property owners. If a feasible plan could be worked out, it might be desirable to exempt improvements on "land," at least in part, so as to remove impediments to the construction of buildings and other improvements. *Fourth*, there would be those special taxes that are clearly justifiable on the benefit principle. In this group would be three major levies: (1) special assessments, (2) highway taxes (on fuel and on vehicles) with the revenue to be spent on highways and streets, and (3) pay-roll taxes *on the worker* (not on the employer) with the revenue to be spent on social security. All three of these have strong claim to use as benefit levies.

This "model" system is as instructive for what it leaves out as for what it includes. There is no place for general sales taxes, for most excises, for taxes bearing especially heavily on risky investments, for pay-roll taxes on the employer, or for personal property taxes; taxes now paid through these levies would be collected instead through direct, personal taxation, geared to ability to pay and unobscured by traditional

⁹ An additional special corporation tax might be imposed on the ground that the government grants special privileges to corporations and is entitled to recoup at least part of the fruits of these privileges.

misapprehensions as to the relative real burdens on different income groups. The reasons produced by the preceding consideration of those particular taxes ought to be sufficient to explain their exclusion. There may be a place in the tax system for fees, licenses, or excises with the specific purpose of controlling the production or sale of "undesirable" goods and services, but such taxes ought to be levied with extreme care lest the actual results be very different from those intended, as in the case of existing tobacco and liquor taxes. At any rate, such levies are not properly classified as major elements in a revenue system, since their major purpose is regulation, not revenue.

A "model" tax system would differ drastically from the present system not only in make-up but also in administration. Experience has shown that centralized collection of income taxes is not only more efficient for the government but much easier on the taxpayer, since he has then to file only one set of returns. Therefore, it might well be desirable to have only one overall income tax with a basic set of federal rates, and state rates added where desired by the states, all collected by the federal government. The funds from the basic levy could be partially retained by the federal government and partially returned to the states; probably in proportion to the amount collected in each state. Any state that had added a special state levy would of course receive the entirety of this levy. Such a program would not only afford efficient collection and minimum bother for the taxpayer, but it also could avoid present undesirable double taxation of income by having all income taxed just once, either at the source or at the domicile of the recipient. A similar collection and sharing system might be used for death, gift, and corporation income taxes, for substantially the same reasons. There is no reason why the governmental unit that spends funds should be the most efficient unit to collect them, though this tacit assumption is apparent everywhere in our fiscal system. General fiscal arguments in favor of greater centralization are also important, and will be considered in the following chapter on fiscal policy and business stabilization.

More efficient administration of the property tax might also be attained, either through state administration of the tax or through closer state supervision to assure more efficient administration by local units. Local assessment and collection procedures on real property are inefficient in large part simply because the local units are so small as to be unable to afford well-trained, efficient assessors and collectors. Probably the solution is state assessment and collection, with each local unit levying that rate that it needs to obtain wanted revenue, with the proceeds redistributed to the local units. The degree of centralization that should be sought in tax administration depends largely on political considerations, but the administrative advantages of more professional methods are indisputably large.

CHAPTER 50

Fiscal Policy and the National Income

TRADITIONALLY the study of public finance has been carried on almost entirely on the tacit assumption of full employment of productive resources. In the last decade, however, a tremendous amount of attention has been given the part that public finance (fiscal policy) can play in raising and stabilizing the level of national income and employment. The preceding three chapters have been devoted primarily to considering public revenue and expenditure policies under conditions of relatively full employment. Fiscal policy in booms and depressions was discussed at considerable length in the earlier chapter entitled "Economic Policy and Business Fluctuations." It is the purpose of the present chapter to integrate these two approaches to fiscal policy. Each is important, and there is no reason why one should be considered to the exclusion of the other.

There is no clear-cut dividing line between the two fundamental cases of "full employment" and "unemployment"—indeed, it is this difficulty that is one of the greatest in formulating economic policies to combat depression and stabilize business activity at a high level. It simply is not possible intelligently to formulate one clear-cut set of fiscal policies for periods of full employment and another for periods of unemployment, and shift sharply from one to the other as "full employment" is reached or lost. In any period of rapidly rising employment and income, full employment and the practical limits of plant capacity are reached in some industries long before they are even approached in most others. Drastic price rises occur at bottlenecks while widespread unemployment continues. "Full employment" has no concise practicably useful definition; but, more important, even if it had, the condition could not be used as an automatic guide to fiscal and monetary action. To continue expansionary financing until every employable person was at work would lead to drastic and widespread inflation. Yet, if continuous care is taken to bear these facts in mind, for pedagogical purposes it is useful to generalize conditions into the two broad cases of full employment and unemployment. In the former—full employment—the function of public finance (obtaining and spending funds) is fundamentally to *transfer* income and resources. After taxes are paid taxpayers can buy fewer private goods, and resources are transferred to production of highways, schools, and so on. If the government obtains the funds without reducing

private spending correspondingly, the result is an increase in total spending and, since full employment has already been attained, higher prices. In the latter case—widespread unemployment—the function of fiscal policy may still be partially the transfer of incomes and productive resources from private to public uses, but it is likewise a major aim to *increase* national income and employment. Thus it is desirable to use financing methods that increase the total amount of spending and income so as to draw idle resources into use. It is along the lines of these two broad conditions that fiscal policy is considered in the following sections, which deal with the use of different types of financing and spending methods under different conditions.

It is appropriate to emphasize again the weaknesses of expansionary fiscal policy alone as a device for increasing employment and national income. When, in the following sections, certain fiscal policies are classified as especially appropriate for periods of widespread unemployment, this by no means implies that their adoption alone will increase employment and income or will even prevent the situation from becoming worse. It means only that, given other appropriate policies, such fiscal steps ought to conduce toward the desired ends. All the material of this chapter assumes knowledge of the earlier chapters on general business fluctuations and on economic policy toward them, and these earlier chapters should be reviewed unless they are well in mind.

Revenues

Taxation

1. **Taxation in periods of relatively full employment.** Taxation is the most direct and precise means of allocating the burden of financing public expenditures. Although the incidence of many taxes is uncertain, for several of the “best” taxes we are tolerably sure where the incidence is, and by using them the burden can be placed roughly where we want it. Thus, when there is relatively full employment and little or no need for expansionary financing, taxes appear clearly to be the best method of financing government expenditures. In such periods government spending involves fundamentally a transfer of productive resources from private to public uses—the primary consideration in raising funds is that the persons giving up control over the use of resources shall be chosen in the most equitable and least disturbing manner possible. Of course the use of taxes rather than borrowing or new-money financing does not assure this end. Some taxes are highly inequitable, some press heavily upon business activity and the level of employment. But in general taxes tend to draw on current income of the taxpayers, thus reducing consumer outlays to offset increased government expenditures and thereby minimizing the danger of inflation. Taxes make possible the noninflationary diversion of resources—private plus public spending power remains roughly unchanged.

It is this fact, that taxation does largely draw on current income, that

makes it especially attractive in periods of relatively full employment. Were government expenditures to be financed at such a time without a corresponding reduction in private spending, total spending would be increased and prices would be pushed up, since total output could be little expanded, in view of the full employment condition. It is in wartime that the danger of using nontax methods of financing under full employment has been greatest—and the world's worst price inflations have centered around war finance based on the issue of new money and government borrowing from the banks. When government spending to obtain war materials soars to unprecedented levels a corresponding amount of new taxes is painful and usually politically unfeasible (at least in the eyes of the politicians). Since the total output of goods is set by the physical limits of manpower, materials, and plant capacities, and since government war needs take increasing proportions of available productive capacity, the total amount of resources left for producing civilian goods is strictly limited and probably declining. Expanding consumer incomes under expansionary government spending, prosperity, and full employment, coupled with a stable or reduced volume of available civilian goods, means inflation—a tremendous upward pressure on the prices of the civilian goods available. Only if civilian buying power is reduced as government spending increases can this inflationary pressure be avoided. In war as in peace, taxation is the most direct and precise means of allocating the burden of financing public expenditures.

2. Taxation in periods of widespread unemployment. However, in periods of widespread unemployment when a major aim of government policy is to raise the level of national income and employment, taxation is a much less attractive means of financing government expenditures. In such periods, government spending does not necessarily involve a transfer of resources from private to public use—on the contrary, the hope is that public spending will help draw into employment idle resources and facilitate a self-feeding increase in incomes and employment. Under such circumstances taxation that reduces the volume of civilian spending on either investment or consumption directly opposes the major aim of government policy. On the whole, taxation is clearly a nonexpansionary means of financing government expenditures in depression periods. The strong arguments against the use of most taxes to finance depression spending have been considered in the chapter on "Economic Policy and Business Fluctuations" and need not be repeated here. If taxes are nevertheless used in depression periods, there is special presumption for avoiding those, such as sales and pay-roll taxes, that directly repress consumption and employment, and for eliminating those tax law provisions which tend to penalize the taking of investment risks by equity capital.

Borrowing

1. "Regular" recurrent short-term borrowing. The revenue from most taxes comes in irregular spurts. Income taxes are usually paid

annually, or quarterly. Property taxes are paid at similar intervals. And so it is with many other taxes. But government expenditures are not spaced in this manner; often there are times when funds are badly needed but will not be received until some near future date. In such circumstances, governments commonly resort to borrowing, usually for short periods. This shows up most clearly at local governments in "tax anticipation warrants," which warrants (sold to banks or other investors) are promises to repay the face value of the warrant at a set future date on which it is anticipated that the taxes will have been collected. Thus the burden of financing government expenditures is temporarily imposed on those who buy the warrants. When the taxes are collected and the warrants paid off, the lenders are repaid and eliminated from the picture (having gained their interest payments) and the burden is shifted to the taxpayers. In such cases little cognizance need be taken of the borrowing, since it is only a temporary deferment of the burden on the taxpayers. At times, when governments fail to collect their taxes and the warrants are not paid off, the burden of financing is placed permanently on the buyers of the warrants. Similar methods of short-term borrowing to tide over temporary shortages are used by state and federal governments, but in the case of the federal government repayment of the short-term "bills" sold is not made contingent upon collection of a particular tax.

But such "regular" short-period borrowings do not bulk large in the total of government borrowing. Major borrowings come when there are extraordinarily large expenditures and when tax revenues are not easily pushed to meet these great demands on the government. Such situations usually arise in connection with wars or depressions, and it is to these "extraordinary" borrowings that primary attention must be devoted.

2. Borrowing in periods of relatively full employment. The urgencies of war finance have been responsible for most cases of large-scale public borrowing in periods of relatively full employment. There has long been lip service to the proposition that taxation is the best way to finance a war, but in the critical hour this principle has generally given way to the easier alternative of borrowing. The present war period is no exception to the rule.

Borrowing to finance government expenditures in war (full employment) periods is not necessarily inflationary. If the borrowed funds draw on current income—that is, if they reduce private spending below what it otherwise would have been—borrowing is no more inflationary than taxation, since spending power is simply diverted from private to public hands. But if the borrowing "draws into circulation" funds that otherwise would not have been spent or involves the creation of new bank credit through sales of securities to the banks, the result is clearly inflationary, since, other things being equal, the total volume of spending is correspondingly increased, while total output cannot be expanded. The government obtains more to spend, while private persons and businesses have just as much as before. The result is a forcing up of prices.

In the present war in this country, much emphasis is being placed on

the sale of War Savings Bonds as a means of averting inflation. To the extent that these bonds reduce private spending on other things they do provide a relatively noninflationary method of financing. Insofar, however, as they are bought out of funds that would otherwise not have been spent (on consumption or investment) they are definitely more inflationary than most taxation. This holds true whether the bonds are bought out of accumulated idle savings or current savings that otherwise would have been held idle. It is simply not correct to assert that "borrowing from the public" is necessarily a noninflationary method of war finance—it may be, but the chances are greater that it will not, especially if the purchases are voluntary.

It is true, however, that if purchases of securities were put on a compulsory basis a noninflationary result quite similar to that of taxation might be obtained. A proposal for such a plan has been popularized under the name of the "Keynes plan," after the noted English economist who suggested it. Under it, part of every person's pay would be withheld each payday and "saved" for him by the government, thereby reducing current private spending power and building up private savings for the postwar period, instead of simply taking the funds once and for all as under taxation. The plan has many variations—under some the withheld savings are necessarily invested in government securities, under others they may be held in the form of "frozen" bank deposits or postal savings accounts. If they are not invested in government securities, the result is to reduce private buying power and check the upward pressure on prices correspondingly, but the government still has the problem of raising funds to finance its expenditures. Essentially such a compulsory savings scheme is similar to taxation, except that it provides for a release of frozen savings, or acts as an "excuse" for government spending, in order to release purchasing power after the war when such action will probably be desirable in facilitating the transfer of resources back to peacetime occupations. However, to obtain any desired war-period reduction in civilian spending, substantially more compulsory savings must be collected than would be required in taxes, because people would look upon compulsory savings as a partial or complete substitute for other savings, which would thereby be freed for spending. By contrast, taxes paid are gone forever, and such payments carry no inducement to make a corresponding reduction in accumulated or current voluntary savings.

If war borrowing from the public on a voluntary basis is likely to be inflationary, borrowing from the banks is certain to be. When bonds are sold to the banks (or to persons financed by bank credit extended for the purpose), new deposits are created for the government to spend. There is no reduction in the spending power of the public. Bank borrowing is thus even more dangerous than borrowing out of idle private savings, because the new bank deposits constitute a semipermanent addition to the supply of "money" beyond increasing the total volume of spending at the time. Yet financing through the banks is the easiest alternative, since it involves only selling earning assets to the banks, which can easily be given

the reserves necessary for such purchases; and it is a course that has been followed almost everywhere.

One other major source of war borrowing remains—the central bank, in this country the Federal Reserve. Direct reliance on central banks for war finance is widely viewed with disfavor, especially in banking and business circles; that it is inflationary cannot be denied, since new funds are created for the government without any reduction in the public's spending power. This firmly established rule of "conservative finance" has its roots in English experience before and during the Napoleonic Wars, when the central bank was viewed as the primary independent check on "unwise" government financial policies. More recently, the central European runaway inflations of the early 1920's were fed by central bank printing-press money when taxes proved completely inadequate to finance government needs. Yet, for any given volume of borrowing, reliance on credit creation by the Federal Reserve is no more, nor less, inflationary than reliance on credit creation by the commercial banks.¹ In either case, the given amount of credit is created and spent by the Treasury. If central bank "greenback" or credit creation is worse than the orthodox method, therefore, it must be because the government finds borrowing easier from the Reserve and therefore places less reliance on taxes and public savings than would otherwise be true. However, under present circumstances in the United States this would probably not be the case, since the Federal Reserve has publicly announced its intention of providing all bank reserves needed for the banks to absorb any securities that the Treasury does not sell elsewhere; whether the Treasury borrows direct from Reserve Banks or through the commercial banks, the Reserve authorities have given assurance that all securities not sold to the public will be taken by the banking system. If direct borrowing were to be used, however, commercial bank reserves would be increased as the newly created funds were spent by the government, and steps would have to be taken to check credit expansion on these new reserves unless further inflationary pressure were to be generated. This could be accomplished by direct controls over new bank loans or by raising reserve requirements to absorb the new reserves.

The case for direct sales to the Reserve Banks rests largely on grounds of superior convenience and directness and on the desirability of having the postwar debt held by the Reserve Banks rather than by commercial banks (to minimize problems of debt and interest payments). The case against the direct sales method rests on the traditional objection that the central bank as an independent agency should exercise a check against excessive Treasury borrowing. In view of the public agreement between Reserve and Treasury authorities to facilitate all desired commercial bank financing at low rates, there appears to be little reason why Federal

¹ Assuming that steps are taken to prevent further credit expansion of the new reserves provided by Reserve Bank security purchases. This point is discussed below.

Reserve opposition to inflationary borrowing should be any different whichever method is used.

Whatever the method used, *to the extent that government finance in wartime or other periods of full employment brings higher prices, it is simply an indirect form of taxation.* The burden of the tax falls on those who lose from the depreciated value of money—bondholders, salary earners, all persons and institutions that have sums due them relatively fixed in money terms, a category that includes large groups of the low income classes. It is a kind of taxation that makes no pretense of conforming to either the benefit or ability-to-pay principles—it meets none of our criteria for a good revenue source. When government war needs soar, consumers simply cannot have as many goods as before. Whether taxes are levied or inflationary financing is used does not alter this basic fact of the limited amount of civilian goods available—it only affects who gets the available goods. Under such circumstances, carefully considered taxation to limit different consumers' buying power is a far more equitable means of allocation than simply letting those who lose most from the depreciated value of money be haphazardly squeezed out. Today, in spite of the tremendous practical difficulties involved, direct government rationing of the scarce goods, rather than either taxation or inflation, is increasingly favored by many persons as the best method of dividing up very scarce goods, in large part because of the strong political resistance to sufficiently heavy taxes to check inflation.

To the equity argument against inflationary financing must be added possible disturbing effects of rapidly rising prices on war production and economic efficiency, once virtually full employment has been reached. This added argument becomes much stronger in the postwar period, when a highly inflated price structure is particularly susceptible to drastic deflation. It is to be hoped that such deflation will not be permitted by fiscal and monetary authorities. But even if this danger is averted, serious post-inflation relative price maladjustments would remain, since different prices and costs rise at such different rates in inflationary periods. On all scores, price inflation in periods of full employment is a thoroughly bad financing method.

3. Borrowing in periods of widespread unemployment. The extent to which government expenditures financed by borrowing are likely to increase employment, production, and income in depression periods has previously been discussed at length.² The overall aim is to increase the total volume of expenditures, and while borrowing and the accumulation of public debt have serious drawbacks at such times, this method is clearly preferable to increased taxation that directly deters production and investment. In principle a strong argument can be made for outright creation of new money or credit by the government or central bank in expansionary financing, though serious practical difficulties must be recognized. Unfortunately, a mere expansion of the volume of bank

² See Chapter 45.

deposits or currency does not insure that there will be a continuing increase in the volume of expenditures. In the long depression of the 1930's, government borrowing helped bring about a tremendous expansion of deposits, but a decline in the rate of turnover of these funds meant a much less than proportional increase in the total volume of expenditures and of national income. The quantity of money and credit in existence is important, but it does not of itself control the level of spending. Unless borrowing from private persons is likely to decrease their spending, it is as good as bank borrowing in depression periods—perhaps better, since the debt is then privately held and an expanded volume of bank deposits is not built up.

4. Long-run effects of borrowing. Much of the controversy over government borrowing arises over its long-run effects on the economy. In particular, two contentions are often advanced: (1) that borrowing simply passes on the burden of financing government expenditures to future generations, and (2) that continued borrowing will sooner or later "bankrupt the economy." In considering these assertions, it is useful to review the policies that may be followed by a government once a large debt has been incurred.

A government debt once incurred may be paid off when due, may be refunded, or may be repudiated. The first case, of repayment when due, is the one commonly assumed when the question of passing on the burden to future generations is considered, but let us dispose of the other possibilities first.

When a bond comes due, the government may simply issue a new security of equal amount to obtain the required funds. This common practice of governments and private corporations is called refunding.³ By this practice the government can avoid ever reducing its total debt. What happens each time refunding occurs is merely a redistribution of income from those who buy the new bonds to those who are paid off, with the government as intermediary. Often, in fact, the group buying the new bonds is almost identical with the group receiving payment for the old ones, especially in the case of banks, insurance companies, and other such institutions that continually maintain large investments in government securities.

How should such a refunding practice be evaluated? The government is fulfilling its obligation to all from whom it has borrowed—there can be no ethical criticism on this score. The major objection is the same one that can be made to any government debt, namely that the continuing interest burden is undesirable since its service may yield unwanted distributional effects or may impede business activity and employment, even

³ The British government often issues "perpetuities" when it borrows—that is, it issues bonds which have no maturity date. These bonds have value simply because of the credit of the government, because of the interest they earn, and because they can always be sold if one wants to liquidate his investment. The same practice is followed by many foreign corporations. A glance at any financial page will show Canadian railroad bonds of this sort.

though only an income transfer is involved. A further objection may be the psychological deterrent on public confidence of a continued public debt, but this is likely to become progressively less important as people get more and more used to an existing level of debt. If the government were ever forced to repay the debt on short notice, the effects might be serious, but so long as there is little likelihood of this necessity, the major objections to continued refunding are the interest burden⁴ and the high existing level when further borrowing becomes desirable. Debt becomes acutely dangerous only when it must be repaid.

Instead of either paying off the debt or refunding it, the government may simply repudiate its obligation. Few persons would approve of the ethics of such a practice, and if the government were to take this action its credit for future borrowings would be seriously impaired. If such action is taken the holders of government bonds have permanently lost the purchasing power given up in buying the bonds. Their situation would be analogous to that of holders of defaulted tax anticipation warrants.

The third alternative is to pay off the bonds when due. This can be done either by taxing to raise the necessary funds or by issue of new money. Suppose taxes are levied. These taxes force taxpayers to give up purchasing power, which is transferred to holders of the maturing bonds. The net effect on total expenditures may be zero, and total employment, production, and income be unaffected, although a deterrent effect on business activity is probable if the taxes used restrict production or if funds are drawn from active sources and paid out to bondholders who hold the funds idle. If the bonds are paid off by issuing new money, insofar as the new money raises prices the transfer of purchasing power is from all who lose by higher prices to bondholders receiving the new funds. If the new money does not raise prices but instead draws unemployed resources into employment, there is no burden on anyone from the issue of the new money even though bondholders receive their principal just as if the funds had been raised by taxes. This is merely a restatement of the effects of financing public expenditures from different sources under varying conditions.

1. These observations indicate the fallacy in the argument that by borrowing we pass on to future generations the cost of present wars or other expenditures. Borrowing is *not* simply postponed taxation. We *cannot* escape the burden of government expenditures by borrowing rather than by taxing, although the immediate and later effects of financing by the two methods may be very different, as already noted. War finance provides a timely example.

The real economic cost of a war is the sacrifice of resources used for

⁴ It is for this reason that some economists argue that it would be advantageous to replace the present practice of financing depression deficits by bonds with financing by the issuing of non-interest-bearing debt (money)—assuming that sufficient safeguards could be maintained against discretionary misuse of the power to issue money.

war purposes rather than for satisfying civilian wants. If we use our resources to produce war materials, we cannot use them to produce houses, dresses, automobiles, symphony music, and other things that we could otherwise have had. A further part of the economic cost of war is the disruption of economic life during the war and afterward when the economy tries to shift back to a peacetime basis. However the war is financed—by taxation, by borrowing, or by issue of new money—these real costs are born by the war generation. They obviously cannot be passed on to anyone else, except insofar as wartime disruption and destruction of capital goods and men impoverishes future generations—and this burden on the future is largely unaffected by the type of war financing used except insofar as war inflation brings postwar deflation.

But may not borrowing impose a special burden on future generations, even though it cannot remove the burden from the war generation? The answer is that payment of interest and principal by any future generation is just a transfer or redistribution of income within that generation. If the debt is paid off, the future generation is taxing itself to pay interest and principal to itself. Clearly the redistributive effects may be very important and there may be effects increasing or decreasing employment, production, and income at that time, but there is no necessary reason why this redistribution will decrease the total wealth and income of the future generation. If the debt is refunded, the result again is a redistributive one and indirect transfer effects are minimized. The belief that we can pass on to future generations the cost of present expenditures by financing through borrowing is almost completely fallacious.⁵ But there is likely to be a different distribution of the cost of the war in the war generation and a different postwar income distribution depending on whether war borrowing or taxation is used.

2. Since payment of interest and principal involves merely a redistribution of income within the economy so long as the debt is domestically held, the fear that increasing public debt may "bankrupt the economy" is equally fictitious. But the argument against a steadily increasing government debt has, nevertheless, considerable validity, for these reasons previously stated: (a) The larger the annual interest burden, the greater is the pressure on tax sources each year to raise funds for interest payments. (b) The fear that heavy government debt will prove disastrous, even when this fear is irrational, may operate to deter production and investment, and may also make it more expensive for the government to borrow additional funds. (c) As a matter of practical politics, when the pressure of making expenditures conform to "regular" revenues is removed, wasteful and inefficient expenditures are much more likely to result. A cumulatively growing public debt has serious disadvantages, but these disadvantages have often been overemphasized and seriously misstated.

⁵ If the war generation borrows funds outside the United States, and the future generation has to pay foreign bondholders, then it is justifiable to speak of a burden of payment being placed on the future generation in the United States.

New money

1. New money in periods of relatively full employment. Not only will financing government expenditures by issue of new money or central bank credit in periods of relatively full employment increase government spending without in any way reducing private spending power, but it is likely to lead to *increased* private spending as well. Since few unemployed resources are available to increase production, the primary result is higher prices, which in turn induce more rapid spending to "beat the price rise." *Insofar as it results in higher prices, the issue of new money is simply an indirect form of taxation.* The inflationary danger of new-money financing is increased by the likelihood that new money once spent by the government will find its way into the banks, since the amount of currency that the public holds in such periods is determined largely by considerations of convenience, and reaching the banks will serve as reserves for multiple credit expansion. Thus for each dollar of new money issued, the total means of payment of society may be expanded several dollars under boom conditions favorable to bank loan expansion. To prevent this additional expansion, special steps must be taken to raise bank reserve requirements as new money is issued. In periods of full employment little defense can be found for the inflationary method of new-money financing, except possibly as a slightly better alternative than reliance on commercial banks.

2. New money in periods of widespread unemployment. The highly inflationary qualities of new money financing under full employment are precisely those that constitute its main claims to use in periods of widespread unemployment when a primary aim is the expansion of total spending. *Insofar as spending financed by new money simply increases employment, production, and income, rather than raises prices, it is a "burdenless" form of financing.* Under those circumstances, no person or group bears any burden of financing the expenditures (for there is no depreciation in the value of money and no taxation), and the purchasing power of some persons and groups is increased by the added income resulting from the government expenditures. In addition to this general "expansionary" advantage of new money-financing in depression, the method has the special advantage over borrowing that it generates no interest-bearing public debt. On the other hand, undoubtedly strong psychological drawbacks exist to financing by the issue of new money or central bank credit, even though it is in essence similar to commercial bank borrowing. These psychological barriers may lead to fear and uncertainty, offsetting any expansion of employment and income that would otherwise have been obtained.

Some persons who favor new-money financing do so on the general principle that continued inflation is what the economic system needs to function smoothly, and that by continually pumping in new money we can cure all our economic ills. Others, however, argue less naively that just as it is logical to avoid expansionary financing under relatively full

employment, so it is logical to use highly expansionary financing under conditions of widespread unemployment. This reasoning points to taxation as the most equitable and efficient financing method for full employment periods, and to government or central bank issue of new funds as the most equitable and efficient expansionary depression financing method. New funds would be issued to aid in raising the level of income and employment. Once relatively full employment had been attained, expansionary financing would be evidenced by markedly rising prices, which would provide a signal for the transition to nonexpansionary taxation and if necessary to deflationary pressure; any necessary deflationary pressure could be exerted simply by withholding tax receipts from the income stream. To answer the objection that the power of new-money financing, once obtained, would be abused and thereby in fact mean price inflation, some advocates suggest simply that some given price level index be taken as a guide to monetary policy—when prices fell below this level, expansionary financing would be used, and vice versa when they rose above it. While such a rule would provide only a crude guide to fiscal and monetary policy, the certainty thus obtained and the avoidance of dependence on arbitrary personal judgments of government officials appeal strongly to persons who favor government by “rules” instead of by discretionary “authorities.” Others would answer the objection instead by saying that if we cannot trust the monetary and fiscal authorities to act intelligently and efficiently, the thing to do is obtain authorities who can be trusted.

Allocations, individual price-fixing and rationing

When production and income are increasing at a rapid rate, as during the armament boom of 1940–41, acute shortages of materials, facilities, and manpower arise in particular industries long before anything approaching general full employment has been reached. For example, acute shortages developed in 1940 in machine tools and men and facilities to make them, in aluminum, in magnesium, and so on. Prices skyrocketed in such cases while millions were still unemployed and plants lay idle. Since an expansionary fiscal policy still seemed desirable in view of the widespread unemployed resources still available, the control chosen was to place individual price ceilings on these particular scarce goods.

As the defense program grew by leaps and bounds, war production needs absorbed virtually all the output of many crucial materials, and the government had to see that manufacturers of defense necessities obtained needed materials. Thus priorities were resorted to, to ensure that scarce materials would be available to defense producers in preference to producers of nonessential civilian goods. A high government-granted priority rating gave a producer preference over other users in buying the available supplies, so that price competition for the supplies was largely eliminated, price ceilings having been set on most of such scarce commodities. When the priorities system was found to be inefficient and incapable of ensuring an effective control over the scarce materials, direct allocations of such materials by the government were resorted to. Pro-

ducers of such scarce materials were directed to supply certain amounts to the various users designated by the government. No one could be served except by government order. The economists and technical industrial experts of the government were gradually taking over the traditional function of the price system in allocating resources.

That the government "experts" made drastic mistakes in many cases is undeniable—but whether adherence to allocation of resources by the price system would have done better in expediting war production is another question. Critics of the policies followed argue that it would have been preferable to prevent general inflation by adopting an aggressive tax policy to hold down consumer buying power to the reduced level of civilian goods available, while leaving relatively free play to individual prices to allocate resources. Advocates of the government allocation policy argued that direct allocation of scarce resources was more certain and quick, that free individual prices could no longer serve their usual function in war time, and that individual price ceilings could be counted on to play a major role against inflation.

The result of the pressures of war and of political considerations, which made very rapid increases in taxes difficult, was a general system of direct controls, constituting widespread priorities and allocations of scarce materials, plus price ceilings on such materials and on an increasing number of the consumers goods made from them. It was widely held that by fixing legal ceilings on individual prices that showed tendencies to rise rapidly the danger of inflation could be averted, though most economists also stressed the need for a strong anti-inflationary tax policy.

Nevertheless, tax increases were far too small to check the inflationary pressure. Without an aggressive fiscal policy to hold down rapidly rising spendable consumer incomes to the reduced level of consumer goods available, the upward pressure on scarce consumer goods prices increased tremendously. National income rose by leaps and bounds, while total goods available to consumers decreased gradually as war production used more and more resources. More and more commodities were put under legal price ceilings. In this setting it finally was generally recognized that under such arrangements prices could no longer act as the market arbiter in determining who gets the goods available—that this job had to be taken over by the government unless the allocation was to be left on a haphazard scramble first-come, first-served, basis. Consumer rationing, difficult and unattractive as it is, under such circumstances appeared to be the best way of obtaining a "fair" distribution of the limited goods available. Legislative or administrative price-fixing does not solve the basic inflation problem—it only shifts it to the "real" level of physical distribution of goods. And gradually direct rationing of scarce consumer goods has been added to the system of government direct controls, thus placing the entire productive process from producer to consumer in such cases under the discretion of government authorities.

In this chapter, direct allocations, price-fixing, and consumer rationing are considered only because in wartime they constitute another course

besides a strong fiscal policy to check inflation. To the extent that a strong tax policy is adopted, the need for individual price-fixing and consumer rationing is minimized; to the extent that consumer spending power is allowed to rise above the level of goods available, price-fixing and rationing become necessary if inflation is to be avoided and goods workably allocated to consumers. It is yet too early to determine which method of approach is the better; probably there will never be evidence to support any clear-cut decision, especially since both are being used simultaneously. That an administered economy is terribly inefficient and inconvenient has been clearly demonstrated, but whether a freer economy would have obtained the all-important war goods faster than has actually been the case is by no means clear. It is, however, significant that price control and rationing authorities have increasingly urged the need for drastic fiscal action if the entire system of direct controls is to be kept from "exploding" under the tremendous inflationary pressure being generated by swollen war incomes. Clearly both fiscal action and direct controls have important roles to play.

Expenditures

Expenditures in periods of relatively full employment

The problems of "regular" government expenditures in periods of relatively full employment, which were considered in a preceding chapter, need little further discussion from the point of view of fiscal policy. Government expenditures increase the incomes of the recipients correspondingly. With full employment they are therefore likely to lead to increased upward pressure on prices—indeed they do so directly when the government itself goes into the market for goods and services—unless the revenues have been obtained so as to reduce private spending power. However, the direction of the public expenditure is of some importance in this connection. For example, payments on bond interest and principal to high income group bondholders are probably less likely to increase the recipients' total spending by the full amount than if the payments had been to low income groups. But in prosperous, full employment periods even this distinction must not be emphasized too much, because neither rich nor poor are likely to hold income "idle." The difference is likely to come instead in the proportion spent on consumption and investment at the different income levels.

Once a period of relatively full employment has been reached, momentum is usually under way toward a full-fledged boom. Under such circumstances it is ordinarily desirable for the government to curtail any current expansionary spending program. Additional taxes should be imposed and expenditures reduced, to ease off the upward swing of the cycle toward stabilization at the high level attained. The principle is simple—it is the problem of timing and execution that is difficult. If the boom is centered largely in certain heavy investment areas, construction for example, obviously government expenditures in these areas should be

especially curtailed. Considerable flexibility in government spending programs is always desirable so that expenditures can be directed where they appear to do the most good and can be curtailed rapidly where they become undesirable. Fixed debt service on outstanding securities is from this point of view particularly unfortunate when the government debt reaches such astronomical levels as in recent years.

The whole fiscal approach in such periods depends largely on how prosperity is to be viewed—as a fragile thing that will collapse back into underemployment if government spending support is withdrawn, or as a situation where the real danger is that speculative overinvestment and inflation will develop, followed by a crisis and ensuing deflation. If the first view is taken, extreme care will be used in decreasing government expansionary spending, especially in the investment industries where it is feared private investment will lag. If the second is accepted, steps will be taken to check expansion and possibly even to apply deflationary pressures before the boom goes to such excesses as to precipitate a crash. During the 'thirties many economists pronounced the economy "mature" and prescribed permanent large-scale public investment if reasonably full employment were to be achieved and maintained. Others demurred equally strongly. The postwar period will do much to show which was nearer right.

In modern war periods the case for drastic anti-inflationary action commonly is clear-cut. The pressures of vastly increased government expenditures and similarly increased private spending upon the limited supply of goods available are certain to be highly inflationary unless drastic preventive steps are taken. On the expenditures side comparatively little can be done. The extent of war spending is dictated by the needs of the armed forces. Nonwar expenditures should be reduced as much as possible, but this reduction is ordinarily so small relative to expanded war expenditures that it is of only minor importance. Therefore, anti-inflationary fiscal pressure must come primarily on the revenue side if it is to be effective.

Expenditures in periods of widespread unemployment

Given the depression period need for expansionary spending through direct relief, subsidies, or public works, there is no clear-cut case for the exclusive use of any one of these methods. As previously indicated, probably all three can be advantageously used. Direct relief is at once the cheapest and most effective device to ensure that funds are quickly respent by the recipients. There is little chance that relief clients will hold idle funds that they receive. On the other hand, the effect of a "dole" on morale is very bad, and it fails to maintain skills for re-employment. Public works are clearly better for morale, and they directly stimulate the investment industries that are claimed by many to have the greatest cumulative "pump-priming" effect. Further, expenditures on public works are "bunched" where they are virtually certain to require the employment of additional resources, whereas the spending of relief

funds may be so spread over many products that no industry is required to increase employment much. Lastly, with public works society has something to show for its expenditures afterwards.

Against these advantages, public works are far the most expensive type of relief and lead to a rapidly increasing public debt if financed by borrowing. More important, they are likely to perpetuate exactly those price-cost maladjustments that are in considerable part responsible for the low level of activity in the heavy industries in which public works activity is centered. Flexibility in timing public works is exceedingly difficult to maintain, partly because of the physical nature of the projects once they are begun, and partly because of the logrolling political pressure to which such projects have become so susceptible in this country with each Congressman out to see that the folks back home "get their share." Further, some people feel that public works are often competitive with, rather than stimulative to, private investments. Lastly, it is by no means certain that funds spent on public works will be respent quickly, especially insofar as substantial proportions go to contractors, materials dealers, and other businessmen who may apply the proceeds to paying off debt, building up cash reserves, and so on.

Direct subsidies to producers can directly increase employment and income, and can do so in a manner directly stimulative to private business enterprise. Otherwise much the same comments apply to them as to public works, except that the political pitfalls involved are probably even greater in direct subsidies to individual private firms than in public works. Traditionally, American public opinion has been antagonistic to such subsidies, in spite of the fact that we make many in slightly hidden forms.

Many economists and probably the bulk of the public favor placing primary emphasis on public works, though for a wide diversity of reasons. It is therefore worth while to consider further certain aspects of public works spending.

"Regular" government expenditures in normal times ordinarily include a substantial volume of public works—highways, parks, bridges, public buildings, and so on. During such periods, that these public works are undertaken presumably indicates a belief that general welfare is increased by having the government take funds from private persons and spend them for the public uses; presumably the marginal loss of satisfaction to the taxpayers is less than the marginal gain to the users of the public works. This is the case insofar as citizens' sovereignty is effective.

Depression public works should be considered in the light of these "regular" needs. Many economists favor "saving up" such public works for depression periods, instead of inaugurating them during prosperity. Others, who accept the "stagnant economy" thesis, feel that even in prosperity a large volume of government public works must be maintained. In either case, it is clear that such "net social gain" public works, if available, should be the ones undertaken first in depression

periods unless there are strong reasons for contrary action. Next, many public works that were marginal in periods of full employment become clearly socially advantageous when the need for expansionary spending arises. For example, a new highway that could not quite be justified on its own merits might become clearly desirable if as a public works project it provided relief for unemployed and aided in raising the level of income and private employment. If the depression is long and severe, and if great faith is put in public works spending, all such projects may be exhausted and it may become necessary to resort to projects of little intrinsic merit. This danger is lessened by careful advance planning so that a useful spending program can be rapidly yet carefully undertaken when the need arises. In all cases special care must be taken lest the projects actually have a net social disadvantage because of the high cost of maintenance after completion.

There are other important criteria in choosing depression public works projects: (1) If possible, the projects should be where the greatest stimulative effect on employment and income will be obtained. As a practical political matter this largely means the construction industry, though differences between its various branches are important. (2) If possible, the projects should be in industries where they will not perpetuate price-cost maladjustments, and should be administered so as to avoid this effect where undertaken. This criterion is likely generally to be in conflict with the preceding one. (3) If possible, projects should be chosen that are relatively flexible so that speed of completion can be altered according to fiscal needs. Fairly short term projects are preferable to long ones on this ground. (4) If possible, projects should be selected that are not competitive with private investment. Insofar as public works replace rather than supplement private outlays no net advantage is obtained.

Integration of Federal-State-Local Fiscal Activities

The fiscal activities of our various levels of government show little integration and often work at cross purposes. This constitutes a serious source of inefficiency in "regular" activities and may completely negate the good effects of positive anticyclical policies adopted by the federal government. In both revenue and expenditure policies the case for better coordination is very strong.

The history of the depression of the 'thirties indicates clearly the dangers involved in such uncoordinated activities. Early in the depression the federal government sharply increased its expenditures to provide relief for the unemployed, financing these expenditures largely by borrowed funds. However, at the same time decreased tax revenues and difficulties in borrowing forced local and many state governments to reduce drastically their expenditures on regular functions. Thus federal spending was largely neutralized—decreases in state and local expenditures more than offset the federal effort. Probably not until late 1934 or 1935

did the increased federal expenditures substantially overbalance state and local contractions. Moreover, a majority of the states resorted to highly deflationary sales taxes and widespread excises to finance their governmental functions, pointing proudly to more nearly balanced budgets. More recently, since the United States entered the war, certain state and local governments have reduced taxes "to make it easier to pay the new federal taxes," when a primary purpose of the federal taxes has been to reduce the funds available to consumers for spending on the limited goods available. If federal fiscal policy is to be effective, the activities of other governmental units must at least be not directly counteracting and preferably should be designed to coordinate.

It is true, of course, that the primary responsibility for anticyclical fiscal policy must lie with the federal government. State and local governments have not the power or breadth to deal adequately with such nationwide problems as depression and inflation. But many economists argue that by a system of federal grants-in-aid to state and local governments state-local expenditures could be maintained through depression periods; expansionary pressures in boom periods could be reduced by shutting off the grants. Such federal grants could be earmarked for specific expenditures in keeping with the fiscal policies adopted, thus utilizing the existing governmental structure to direct stimulative payments.

The case for integration is similarly strong on grounds of efficiency in "normal" periods, as to both revenues and expenditures. As indicated in the earlier discussion of a "model tax system," experience has shown clearly the great advantages to be had from centralized collection of income taxes and centralized administration of local property taxes, both supplemented by revenue-sharing arrangements to return part of the funds to subsidiary governmental units. On the expenditure side, parallel advantages of integration appear. It is apparent that many areas of the country are so poor that they are unable to provide educational facilities, roads, police, fire and health protection, and other such "essentials" at levels consistent with minimum American standards. This poverty becomes even more striking in depression periods. Other governmental units, in spite of adequate financial ability, fail to maintain decent standards because of shockingly inefficient administration. To the extent that national minimum standards in education, public health, highways, and so on are to be maintained, such areas must be subsidized by the rest of the country or local governments required to meet higher standards. One important way of achieving this result is through grants-in-aid from superior to local governments, conditioned on the requirement that the minimum standards be met—an arrangement which ensures the maximum of state-local spending autonomy consistent with minimum performance standards. Such grants have been used on a large scale by the federal government for many years, especially in the fields of education, highways, and public health. Their potentialities extend, however, to all fields where it is desirable to establish minimum standards in

particular areas, ranging from relief in depression to such permanent projects as soil conservation and low-cost housing.

Greater centralization of control would probably result in greatly increased efficiency in both the revenue and expenditure systems, and establishment of minimum standards would go far toward eliminating serious gaps in the public services provided in numerous areas. On the other hand, however, the arguments in favor of local self-government and "states' rights" have much merit, and these considerations must be weighed against the advantages of greater centralization. In any case, the political barriers to such increasing centralization are great, and state-local assent to drastically increased federal powers can probably be obtained only if such assent is necessary to obtain proffered federal funds. On whatever governmental level the powers lie, in the last analysis wise planning and efficient administration of public revenue and expenditure policies depend on an intelligent, aroused public interest and on the securing of capable, honest public officials.

CHAPTER 51

Social Security: A Case Study

MODERN industrialization, and especially the events of the 1930's, has brought all groups face to face with problems of economic insecurity. Industrial workers, farmers, clerks, businessmen, professional men, and large-scale investors, even before the impact of the present war had become uncertain of their security to an extent unknown in the United States since earliest colonial times. During the long depression of the 'thirties, two of these groups—wage earners and farmers—perhaps more than any others bore the brunt of this hazard, which many social scientists have called the greatest bane of modern civilization. Today the war and the administered economic life which it involves have completely reversed the circumstances of the preceding decade. Yet there is, unfortunately, much reason to suppose that with the peace we shall again be faced with the same fundamental problems of economic insecurity that loomed so large in past decades. Social security activities will thus constitute a large and probably growing segment of the public economy. This chapter, largely disregarding the special circumstances of wartime, deals primarily with the problem of social security for wage earners and their families, using government action in this particular socio-political-economic sphere as a cross-section study of the public economy in action.

The Background of Social Security

Until recently, in the United States, an inchoate system of local charities was depended upon to shoulder the burden of caring for persons unable to support themselves. Inability to provide self-support was considered a social stigma, and the outspoken general attitude was one of "If he can't make his own way, why should we carry him on our shoulders?" Workers were expected to take jobs at their own risks; special compensation for industrial injuries was almost impossible to collect. The specter of poverty in old age, which so haunts the low income groups, was one of those uncomfortable subjects to be avoided in polite conversation. Although local "poor laws" existed in most communities from early times, they were in most cases so ineffectual as to be only a sop to the consciences of those more fortunately situated.

With the continued spread of industrialization and its attendant problems of urban concentration, abominable factory conditions, and recurrent business depressions, these problems of economic insecurity became

more widespread, more insistent. The need for some form of organized action to deal with the various types of economic insecurity was increasingly obvious to those foresighted few who were willing to face the facts, yet little recognized by most of the middle and upper income groups influential in government policy determination. By the late nineteenth and early twentieth centuries, programs of state-supervised relief for "unemployables" had become fairly general, being applied to such groups as the blind, disabled, and mentally unsound. During this same period, admitted state responsibility for poor relief became more general, though still on a very meager scale. Mothers' pensions and special aid for unsupported children fell into the same category. American social security measures lagged far behind those of England, Germany, Sweden, and other European countries.

It was not until the nationwide economic debacle of the early 1930's brought inescapably to public and governmental attention the basic problems of social insecurity that federal poor relief was undertaken on a significant scale and social insurance was considered a matter for large-scale federal action. Largely under the pressure of political necessity, the federal government took over from faltering private and state-local governmental hands the job of caring for the unemployed and needy. From this makeshift arrangement there gradually emerged a full-fledged two-part federal program—"work relief" and social security legislation—not only aimed at mitigating unemployment due to depression but also designed to provide old-age assistance and aid for continuously needy groups. The early impetus toward special social security measures, reinforced strongly by the Roosevelt Administration's desire to help the "underfed and underprivileged third," culminated in the Social Security Act of 1935. Work relief programs, which had originally existed primarily as necessary means of feeding the unemployed, gradually shifted in emphasis first to "pump-priming" and then to "compensatory spending" aimed at combating a supposed permanent state of economic "maturity" or "stagnation." At the outbreak of the Second World War in 1939, the social security program in the broader sense represented primarily large-scale public works spending and a widespread program of special social security measures, centralized to a considerable extent in the federal government but operating unequally on all three governmental levels.

The diverse moving forces behind the two parts of the Administration's social security program led to differing and sometimes even contradictory programs in the "work relief" and narrower social security spheres. While the work relief program gradually developed from a pure political-economic expedient into a more or less coherent spend-for-prosperity program, the narrower social security measures rested almost exclusively on "social welfare" considerations and paid little attention to fiscal implications. Behind both lay the basic desire for economic security for the low income groups, at the cost of economic "progress" if need be. Both work relief and the newly-enacted social security meas-

ures were aimed simultaneously at minimizing insecurity and redistributing income in favor of the poor. But in spite of their common aims, the two branches were pushed by different groups with different immediate goals; conflict, especially in the field of fiscal policy, appeared repeatedly, flaring finally into open quarreling in the discussions of war finance during the early 1940's.

That economic security should rank above rapid "economic progress" was a proposition little questioned outside academic discussions after 1933. Strong emphasis, however, must be placed on the haphazard political development of the programs enacted. Seldom do politicians follow closely integrated "programs" advanced by sociologists and economists. Instead law-making is a process of compromises and trading between interests at stake, and social security legislation has been no exception. The number of dollars voted for work relief in depression depends much more on how hungry people are and how likely they are to vote for the other party than on any nice economic calculation as to the amount of pump-priming or compensatory spending necessary to restore full employment. The size of old-age pensions or of unemployment insurance benefits is likewise likely to depend far more on such considerations than on the expert advice of the social workers and economists. Technical professional advice has been used more generously by the Roosevelt Administration than ever before in federal government, but it would still be a major misconception to suppose that Congressional or executive policies on social problems are formulated by weighing carefully the various arguments advanced by professional social scientists.

The nature and scope of the federal work-relief program in the 1930's has been discussed at some length and need not be reviewed here. Minimum wage laws and governmental support of trade-union efforts to improve the lot of the industrial worker, measures that might be considered part of the government's social security system, have likewise been considered in detail in earlier chapters. Therefore this chapter, except in the concluding section, deals primarily with social security provisions in the narrower sense of the term—more specifically with the provisions of the federal Social Security Act of 1935 and of the state laws requiring workmen's compensation plans. The concluding section attempts to view these provisions as a whole in the light of broader social-economic problems and to evaluate briefly in this perspective the economic aspects of the social security program that has evolved rapidly and come to concrete form primarily since 1933.

The Federal Social Security Act

The magnitude of the program covered by the Social Security Act of 1935, as later extended, may be indicated by certain figures for the fiscal year ending June 30, 1940, a period chosen in order to avoid the drastic impact of the war on labor markets and employment condi-

tions beginning in 1941. In fiscal 1940 total federal and state funds paid to individuals under the federal Social Security Act were \$1,085,800,000.¹ This included \$585,700,000 to recipients of such direct aid as old-age pensions and mothers' and children's assistance; \$484,500,000 in unemployment benefits under cooperating state laws; and \$17,600,000 in payments under the old-age insurance system, this last sum representing the small payments for the first half year of old-age insurance disbursements under the Act.² These payments were in addition to over \$1,500,000,000 of federal expenditures for direct relief and public works, to perhaps \$1,000,000,000 of state and local relief and welfare payments not included above, and to private charity. A special federally established security system exists for railroad workers, and somewhat similar provisions are made for other occupational groups (for example, the Civil Service Retirement System for federal employees). Social security is a huge government enterprise, having constituted in 1940 over 20 per cent of all governmental expenditures.

It was estimated that on the average during fiscal 1940 more than 21 million workers were in occupations covered by unemployment compensation laws, and that between 23 and 25 million were covered by federal old-age insurance. The totals of those having unemployment or old-age credits accumulated were larger, amounting to over 40 million for unemployment insurance and nearly 50 million for old-age insurance accounts, since many persons not employed in 1940 had earlier accumulated reserve accounts. The weekly average of workers receiving unemployment benefits exceeded 873,000, with an additional 161,000 under the Railroad Unemployment Insurance Act. In addition, state employment exchanges, financed half by federal contributions matched against state funds and coordinated by the federal Social Security Board, carried through steps in filling some three and a half million jobs. In persons affected as well as in dollars spent, social security represents a far-reaching government enterprise.

• Unemployment Insurance

The legal status of unemployment insurance

England, Germany, and several other European countries began experimenting with various unemployment insurance plans long before serious attention was given the problem in the United States; and, to a considerable extent, the United States arrangements reflect European practices. Although the Wisconsin Unemployment Insurance Act, reflecting long-continued agitation for social reform in that state, became law in 1932, passage of the federal Social Security Act of 1935 represented essentially a pioneer effort toward American unemployment insurance.

¹ *Annual Report of the Social Security Board for 1940.*

² By the end of fiscal 1941 total assets of the old-age insurance trust fund totaled almost \$2,400,000,000, as premiums were collected regularly and initial benefit payments were small.

The section of the Social Security Act dealing with unemployment insurance does not establish a federal insurance system. Instead, it provides for special aid to those states that establish acceptable systems, while penalizing those that do not. The law imposes on employers a federal tax equal to 3 per cent of wages and salaries paid up to \$3,000 per annum to each employee. However, in states having unemployment insurance plans approved by the federal Social Security Board, employers are granted exemption from the federal tax to the full amount of the state tax paid for unemployment insurance up to nine tenths of the federal tax.³ The obvious intention of the law was to induce all states to adopt acceptable insurance plans, and by 1939 all states had qualified under the federal Act.

Important exemptions from the federal tax are provided by the Act. For administrative convenience, a blanket exemption is granted to all employers who employ less than eight workers. In addition, exemptions are granted in the cases of (1) agriculture, (2) domestic service in private homes, (3) federal and state government service, (4) employment of members of the immediate family, (5) navigation enterprises, and (6) nonprofit, charitable, religious, scientific, and educational institutions. While the majority of all employees are covered by the program, this list of exemptions indicates the substantial proportion of the working population excluded under the federal Act, and therefore probably not covered by most state laws.

In order to qualify under the federal Act, state laws must meet essentially the following requirements, although exceptions are permitted under some circumstances. All benefit payments must be made through public employment offices or other state-approved agencies; reserve funds must be deposited with the federal Treasury for investment in government securities; no payments may be made until two years after the inauguration of a state plan; regular reports must be made to the Social Security Board; and benefits must not be denied to any eligible individual for refusing to accept employment where there is a labor dispute, where he is denied the right to join a *bona fide* union, or where the conditions of employment are substantially less favorable than those prevailing for similar work elsewhere in the locality. Except for these requirements, states are left free to enact such provisions as they like, and considerable diversity exists. The purpose of the federal Act is to ensure employees of reasonably satisfactory protection; and where special state or private arrangements meet this standard they are acceptable. For example, some state laws permit employers to take out unemployment insurance for employees with private companies or to establish their own reserve funds. In some cases, a minimum annual wage guarantee or a guarantee of thirty hours work per week for forty calendar weeks per

³ The 10 per cent of the tax collected by the federal government goes for administrative expenses and for varied-purpose grants to the state insurance programs.

year has been substituted for the more common insurance arrangements. Where state laws permit such special arrangements, the employer may claim his 90 per cent credit against the federal tax just as if he were participating in a state unemployment insurance plan. But eventually, probably greater uniformity will be attempted on the basis of comparative records of different state practices.

Is "unemployment insurance" really insurance?

Unemployment insurance is, by the very nature of the unemployment risk, fundamentally different from life, fire, theft, and other regular types of insurance. Ordinary insurance rests on two assumptions: (1) the predictability of the risk involved when a large number of cases are considered, and (2) the strong improbability of a mass occurrence of the insured event. Neither of these conditions necessary for ordinary insurance exists in the field of unemployment.

1. Although the private insurance company cannot tell whether any given building will burn down in any year, it can estimate closely on the basis of experience how many buildings of a particular sort will burn per year of a very large group say fifty thousand. On the basis of this estimate, the company can tell about how much in benefits it will have to pay out per year on a large number of buildings. It can then set an insurance rate on such buildings sufficient to cover the probable payments to be made plus expenses of operation. Because of the actuarial basis of likely fire losses for large groups of insured buildings, any individual owner can convert the danger of an unpredictable large loss into a relatively small annual insurance cost. The same sort of actuarial calculation underlies life, accident, theft, and other such sorts of insurance. By contrast, the total amount of unemployment in any given year, or over any group of years, is highly unpredictable. There is no actuarial basis for predicting the number of workers out of a large group who will be unemployed in any given period. Unemployment depends on many factors—the decline of industries, replacement of workers by machine methods, depressions and the like—and most of these factors themselves are highly unpredictable.

2. Experience has shown that unemployment is likely to come on a mass scale during depression periods. Thus, the unemployment of one worker is likely to be associated with, and partially the cause of, the unemployment of other workers. It is as if all the buildings insured by a fire insurance company were attached, so that if one burned all the rest were thereby endangered. In fire insurance this is termed the "conflagration risk," and no company would think of assuming such a risk on a large scale. To afford to do so, the concern would have to charge extremely high premiums, and unless huge reserves had been accumulated it still might be wiped out financially by a single fire. Instead, by insuring individual buildings in scattered localities or by reinsuring with other companies in case of risk concentration, most insurance companies

hold their payments to actuarial expectations and operate on roughly a current basis, with current income sufficient to cover current payments on claims and the company reserves left in investment.

These considerations indicate the special problems of unemployment "insurance." Unemployment is not predictable with any real accuracy, and it becomes a mass risk not susceptible to the "spreading" that is the essence of regular insurance. Moreover, in the case of ordinary insurance, the insured person can usually contribute to a lower premium rate by reducing the likelihood of fire, accident, or theft in his own case; therefore he has a special incentive to help prevent the event insured against. But since the unemployment problem centers so largely in economy-wide depressions, there is relatively little that the individual employer can do to lessen the likelihood of unemployment for his workers in such periods. And the rapid breakdown of well-established unemployment insurance plans in England and Germany during the depression of the 1930's illustrates the inability of regular premium-accumulated insurance reserves to meet for long the enormous demands of severe depression.

Economic aspects of unemployment insurance

If unemployment insurance is not insurance in the strict sense of the word, what is it? The answer is that fundamentally it amounts to a special and systematic form of unemployment relief, financed in the United States by pay-roll taxes levied on employers. The present plan aims to collect during periods of relatively full employment large sums to be disbursed when unemployment arrives. In essence, the plan might simply provide for "relief" to the unemployed equal to present insurance benefits and finance these relief payments out of special employer taxes without changing anything except the name. Yet there seems to be considerable reason for retaining the name of insurance, primarily on the psychological grounds of removing the stigma of charity from payments made.

Looked at economically, the problem becomes basically one of how large unemployment benefits should be, how the funds should be raised, and how the payments should be distributed. These are precisely the questions discussed at some length when alternative fiscal and monetary policies to mitigate depressions were examined, and what was said there applies equally well here. Payment of large unemployment benefits is, on economic grounds, similar to direct relief; and insofar as the primary aim of depression spending is to restore full employment, the optimum volume of unemployment benefits is the same as that for relief payments, assuming that the two are alternative policies. How much should be spent to stimulate re-employment? Should the funds flow through public works, subsidies, or direct unemployment benefits? Should depression spending be financed by borrowing or taxation to be most effective, and if by taxes by what sort of levies? The answers to these general questions provide, insofar as economic considerations go, the basic answers

to the more specific questions faced in planning unemployment insurance.

During consideration of the Social Security Act, it was suggested that unemployment insurance premiums accumulated in good times should be invested in open market government securities, and that these securities should be liquidated when needed to obtain funds for benefit payments. The perverse effects of such a policy are evident. Large-scale investment of insurance premiums in good times would contribute to security booms and pour investment funds into the market, while the mass liquidation of securities in depression would not only involve partial loss of invested premiums resulting from price declines but would also strongly reinforce the downward spiral of liquidation and contraction, greatly increasing the task of monetary and fiscal authorities in checking the decline.

In practice, it was evident that investing insurance funds in the open market was impracticable, and instead accumulated funds have been invested by the state insurance systems in special Treasury issues that are not salable in the open market. Under this arrangement, the net effect of social security financing depends on the use made of the funds by the Treasury. Insofar as pay-roll tax collections are used to finance current government expenditures, the insurance trust fund becomes merely a nominal "fund," and part of the deflationary effect of the pay-roll taxes is offset. Insofar as the Treasury withholds the funds from the income stream, the pay-roll tax exercises its full deflationary effect and the overall effect is identical with cash accumulation by the insurance fund. Social security technicians have long insisted that unemployment insurance trust funds must be treated purely as insurance reserves and must not be involved in anticyclical fiscal operations. Yet clearly such a separation is impossible. Whether we like it or not, accumulation of unemployment insurance funds inevitably has potent fiscal implications, and in view of this fact there is every reason to adjust social security procedures to aid rather than hinder stabilization efforts.

In view of these considerations, three points may be especially noted in connection with the present unemployment insurance plan:

1. **Coverage.** Since unemployment insurance is fundamentally organized relief rather than "scientific" insurance, there seems to be little ground for continuing to exclude a large portion of the working population. Social Security officials have already argued strongly for extending coverage to all excluded groups except possibly agriculture (because of administrative difficulties), and there appears to be much in favor of this proposal if unemployment "insurance" really has substantial advantages over direct relief. Likewise, the present restriction of benefit payments to twelve consecutive weeks might be removed or loosened. The argument underlying the present rule is that such a restriction is needed to maintain the solvency of the plan, but this solvency is largely technical since the government is forced to assume the relief burden for the unemployed who cease to receive insurance benefits. In view of the serious

drawbacks of the present financing methods noted in the following paragraph, however, extensions of coverage should be made only in the event that the method of collecting funds is changed.

2. Financing. Many persons feel that because unemployment insurance is financed by a tax levied on the employer, the burden is borne by the employer. Such a tax, however, is generally shifted onto the worker or leads directly to unemployment. During depression periods the use of the pay-roll taxes on the employer for any expansionary purpose is self-defeating. In boom times, pay-roll taxes are highly deflationary, which may be desirable, but they are also highly regressive. To make unemployment benefits large enough to provide living incomes for all workers during prolonged periods of unemployment, pay-roll tax rates would have to be greatly increased. This would place the burden of financing relief directly upon the working class, upon those on the whole least able to bear the burden. If, as appears to be the case, national sentiment is in favor of less inequality of income and of helping the lowest income groups, there is much to be said for financing unemployment "insurance" primarily out of general tax revenues obtained through a progressive tax system as a standard policy, with appropriate use of borrowing in depression periods. Ideally, the sources of unemployment insurance payments should be determined on the basis of criteria for sound fiscal and monetary policy, not as a special revenue source geared to particular benefits derived by payment recipients.

The best solution may be to retain unemployment insurance as one part of a broad program to provide benefits for the unemployed, supplementing these benefits when necessary by public works and direct relief. During prosperity the use of insurance premiums would be determined by general fiscal policy. If this procedure should be followed, there is much to be said for asking workers to pay part of the cost of their own insurance (if such it is to be called), but this might better be done through direct taxes on wages than through the present disguised wage-levies, which also act as direct deterrents to employment. The bulk of any necessary relief program could then be financed by other means, primarily borrowing or progressive taxation. In any case, the relief problem must be recognized as one of national scope, to be handled on some more or less uniform basis by national authorities.

3. Benefits. Benefit payments under unemployment insurance, as under direct relief, ought to be substantially below those obtainable in private employment unless this would place the recipient's income under some "minimum subsistence" level. Experience has shown that in many cases relief does become habitual and relief recipients become unwilling to return to work when government payments almost equal those in private employment. Relatively low government payments constitute the strongest safeguard against "loafing on the dole." Once again, however, the question becomes involved in considerations of monetary and fiscal policy concerning the best type of government expenditures to

combat unemployment. Likewise, the sociological and psychological considerations involved in the treatment of unemployed persons are often not in accord with those of economic reasoning. In any case, it is generally agreed that payment of unemployment benefits should be made contingent upon the recipient's registering at a government employment exchange and his accepting employment when offered at "prevailing wages" and under "prevailing working conditions."

Whatever the size of the benefit payments made, it should be recognized that this is not the end of the government's responsibility. In depression the need for training youths and re-training older unemployed workers for useful jobs is tremendous, and this responsibility of government should be correlative with that of feeding and clothing the needy. Government re-training and interim subsidies to aid workers in adjusting to industry shifts could greatly ease the depression suffering of many skilled groups, and the training of maturing youths for useful jobs would both mitigate depression difficulties and build for the future. The great lag of the United States in accepting these governmental responsibilities is inexcusable; many other nations have long made government-financed vocational activities a central part of their social security programs.

Old-Age Insurance and Aid to Special Needy Groups

The Social Security Act provides for a system of old-age insurance and annuities, administered solely by the federal government. Insurance is made compulsory for all workers except those specifically exempted, and, except that firms hiring less than eight workers are included in the old-age plan, the exemptions are roughly the same as for unemployment insurance. Old-age insurance is partly financed by a pay-roll tax similar to that of the unemployment insurance plan, the rate (1942) being 1 per cent of the wage or salary of the employee. For old-age insurance, however, a like amount is deducted directly from the wage of each worker, making a total premium rate of 2 per cent of wages or salary.

In theory the plan is set up to resemble an ordinary insurance company, with premiums (taxes) calculated roughly on an actuarial basis and reserves accumulated to strengthen the fund and assure payments when policyholders' claims come due. The funds thus collected are required by law to be invested in United States government securities of certain types, with the interest received by the Old-Age Insurance Fund from the Treasury being added to accumulated tax receipts to finance benefit payments when they come due. The annuity payments provided by the plan begin at age 65, assuming that the policyholder is willing to give up gainful employment, a necessary condition of receiving payments. The annuity received by any worker bears a proportional relation to the premiums he has paid in (and therefore to the salary received during his working years), but in no case may it exceed \$85 per month. If the insured dies before reaching 65, the monthly payments to which he would have been entitled thereafter are paid to his widow, unmarried children under 18, or

to his parents under certain conditions. If there is no survivor eligible to receive these monthly payments, a lump sum, calculated on the basis of the premiums paid in, is paid to the estate.

Although from the point of view of the insured individual there is a close similarity between this system and ordinary annuity policies sold by private companies, from an overall point of view certain important differences exist. The most crucial of these centers around the vast amount of funds collected and paid out under a nationwide compulsory insurance plan.

In the discussions preceding inauguration of the federal plan, some persons favored accumulating a huge reserve fund of many billions in cash, but the tremendous deflationary pressure of completely withdrawing such a sum from the income stream over a period of years was obviously something to be avoided. Therefore, two alternatives were available. One was to put the plan on a pay-as-you-go basis, with taxes adjusted to approximately equal benefits each year. The other was to levy actuarial old-age rates from the outset, investing accumulated premiums in government securities. Eventually the two plans would become almost identical, when, around 1980, the number of new persons entering the plan will roughly equal the number retiring or dying. But in the meantime use of actuarial rates would pile up large reserves (to around \$47,000,000,000 by 1980) because those becoming eligible for benefits would have relatively small accumulated payments while full premiums were being paid into the reserve fund by all insured persons. On the other hand, a pay-as-you-go policy would have meant very low pay-roll tax (premium) rates at the outset but higher ones eventually than would be necessary if reserve funds were accumulated over the early period.

The alternative chosen was that of accumulating the reserve fund, largely because this was thought to give a fairer distribution over time of the costs of old-age insurance. Moreover, since a huge government debt was already outstanding and growing in the 1930's, social security reserves could be invested in government securities without requiring any increase in the public debt or in debt service charges. Instead, this use of the insurance reserves reduced the amount of "outside" borrowing required to finance current government expenditures. None the less, as might have been expected, the heavy pay-roll tax collections proved highly deflationary, especially in view of the large-scale unemployment during the 'thirties, and probably substantially offset more expansionary federal fiscal policies.

As a result, the plan has in essence been put on a pay-as-you-go basis, with a 1939 amendment requiring the trustees to make a special report to Congress requesting tax changes whenever they think the fund is dangerously reduced or more than three times larger than estimated benefit payments for the coming year. Presumably, if such a report were made, Congress would take steps to increase or decrease tax levies for the fund. This pay-as-you-go basis, used by private insurance companies, appeared

essential in view of the monetary implications of collection and withdrawal of funds on such an unprecedentedly large scale.

Aside from the huge size of the old-age insurance plan and its monetary-fiscal implications, the major difference between it and private insurance lies in the method of collecting premiums. Private insurance is voluntary and individually financed. The public plan is compulsory and tax-financed. The tax deducted from the wages of the worker is obviously an expense to the worker, but, in view of the very moderate cost of the insurance to the individual, a thoroughly justifiable expense. Continuance of the tax on the employer is less warranted. This tax is virtually certain to be shifted to the employee, and is likely to cause unemployment as well. If we want to compel workers to buy old-age insurance, the total tax should be placed directly on them instead of getting essentially the same final burden but additional unemployment problems through use of the employer pay-roll tax. If we want to provide old-age insurance for workers without placing the full cost on them, there is much to be said for subsidizing the insurance system out of general tax receipts. This would place some of the burden on groups other than the workers, although workers would finance part of the subsidy through their own regular tax payments. If the subsidy should be financed by those best "able to pay," the solution is the use of progressive taxes bearing most heavily on the highest income groups.

Old-age pensions

The inadequacy of private charity as a means of support for indigent aged persons has become increasingly evident throughout the last century, and state and local relief funds have gradually been set up to supplement private assistance. But it was not until 1935 that provision was made by the federal government for giving systematic aid to this group. Part of this aid comprises the old-age insurance system just discussed, which provides income in old age for participants regardless of need. A quite separate part is the provision that the federal government will match state pension payments to needy aged persons up to \$15 per individual, making a total matched payment (federal plus state) of \$30 per person. Since the maximum federal contribution is \$15, if a state wishes to increase the total payment over \$30, it must provide state funds to finance the entire excess. Actually, many states have failed to take advantage of the full federal contribution, and total payments in those states are less than \$30 per month. If the aged person has other sources of income, the pension he receives is likely to be reduced from the maximum available.

Considerable agitation has occurred in favor of easing the terms on which federal funds are made available for the aged needy. The extreme poverty of many old persons, even after receiving the combined state-federal grant in many states having low maxima, is pointed out. The gist of many of these suggestions is that the federal government make available funds beyond a dollar-for-dollar basis to states that are unable to provide their full \$15 share, or that the federal government increase

its matching by offering to contribute more than the present \$15 limit. Accompanying these suggestions is often the proposal for closer federal supervision of the standards under which the states provide pensions. Many cases of inefficient and even corrupt handling of state and local pension funds have come to light, and since the distribution of federal funds is carried on through state and local agencies there appears to be good reason why additional federal aid should be made contingent upon improved administration.

Aid to other special groups

The federal government extends grants-in-aid to the states for numerous other purposes. Financially most important of these are the grants to aid in caring for needy dependent children, where the federal government will provide one third of the total benefit payment up to a maximum of \$18 total payment for the first child in a family and \$12 for each other child. Other grants are made to the states to aid in caring for the blind, for extending and improving local maternal and child health programs, for extending rehabilitation programs for the physically disabled, and for extending assistance programs for crippled, homeless, and delinquent children. Further, large sums are granted regularly to provide for more adequate public health services.

Under the Social Security Act, the federal government has also taken steps to coordinate state and local employment exchanges into a national system, both to aid unemployed persons in finding jobs and to aid employers in reaching available labor supplies. Faced with the almost complete absence of effective state or local publicly operated labor exchanges, the progress of the Social Security Board has been slow. However, the national defense program provided a drastic stimulus to the public employment exchanges program, and, coupled with governmental training programs and "refresher courses" for jobs crucial to the war effort, the public employment exchanges have recently played an important part in increasing worker mobility. But in spite of this development, the United States is still far short of equaling the employment exchanges and placement programs publicly operated in many European countries and in Australasia.

Workmen's Compensation

The risk of industrial accident, with its attendant loss of working time, medical expenses and often permanently impaired earning capacity, has long constituted a major sphere of insecurity for wage earners. Since the beginning of large-scale mechanized industrial production, accidents have taken a devastating toll of workers, even into recent years when the use of safety devices has spread rapidly. While government action to alleviate this type of insecurity has been mainly along regulatory lines and falls outside the public economy proper, the legal requirements imposed on employers have in effect constituted an integral part of the social

security program and must be considered in connection with other security measures.

Under the common law, workers could obtain damages from employers in cases of industrial accidents only when the accident could be proved to be the result of the employer's own negligence. In effect this principle barred the employee from collecting damages in the vast majority of cases, even when the employer's responsibility was clear. Not only was the employee usually in no financial condition to fight through a lawsuit, but he knew that his job would vanish if such altercations with his employer were incurred. Moreover, the law gave the employer three special defenses, any one of which could prevent the worker from collecting: (1) the doctrine of "contributory negligence," which provided that if the worker himself was negligent to any degree, however slight, he could not collect damages, no matter how great the negligence of the employer; (2) the "fellow servant" doctrine, which provided that if an accident was the result of another employee's negligence the employer could not be held liable; and (3) the "assumption of risk" doctrine, which held that a worker assumed the "ordinary" risks of the occupation when he took a job, and could not collect for injuries received through assumption of such "ordinary" risks. The definition of ordinary risks depended largely on the courts, and generally the range the employer could get classified as "ordinary" was large. In view of these forbidding legal barriers, coupled with the worker's limited knowledge and financial resources, it is little wonder that (in cases of injuries from industrial accidents) the worker generally either collected nothing or had to be satisfied with a small settlement out of court.

The list of states adopting compulsory workmen's compensation laws to aid the worker has grown steadily but slowly. There is no comprehensive federal law on this subject,⁴ and the state laws in force vary widely in their provisions. The core of almost all workmen's compensation acts is, however, the provision that the employer is responsible for injuries resulting from any accident "arising out of or in the course of employment." The passage of such an act replaces the provisions of the old common law, and the worker need not prove employer negligence to collect for accidents occurring on the job. In view of the obvious intent of such laws, the courts have generally been strict in holding employers responsible when cases have been brought into court. Under most state laws, employers are required to take out accident insurance with approved companies so that employees' claims will be fully protected. Some states have set up their own insurance funds, to which employers must contribute premiums on a regular insurance basis.

Workmen's compensation laws mark an important step in the broadening recognition of the need for public action to assure social security for

⁴ Special provisions are made for federal and District of Columbia employees and for marine personnel at sea

the lower income groups. Loss of working time through accident may not be serious for a man making \$10,000 a year, but it is a constant specter for one earning \$1,000. Moreover, the high income employee is often protected by the terms of his contract or the customs of his occupation; he is usually hired on an annual basis, and adequate allowance for sick leave with pay is generally provided. Furthermore, the risks of industrial accident are low for most upper income jobs. These factors being recognized, the basic shift responsible for the spread of workmen's compensation acts has probably been the growing ethical presupposition that society has a responsibility to protect the employee from injustice in case of accident. Were their incomes higher, wage earners might be expected to save enough to protect themselves against the "rainy day" when they are laid up by an accident; but the great mass of workers are unable to make such provision, or even to purchase individual accident insurance policies.

Supplementing this ethical argument for workmen's compensation are others. One of the most important of these is the fact that compensation acts have proved to be highly effective in reducing the number of accidents. When employers become liable for damages arising out of industrial accidents, they have a strong incentive to install safety devices and take all possible precautions to prevent the occurrence of accidents. The remarkable safety records of recent years, as compared with the shocking accident rates of the past century, are in considerable part attributable to workmen's compensation provisions, under which an employer's insurance premiums often vary inversely with his accident rate.

Another argument widely advanced in support of workmen's compensation is that injuries to workmen are as much a cost to be assumed by the employer as is wear and tear on buildings and on machinery. In spite of this argument, however, it is probable that in most cases the cost of accident insurance is largely shifted back to the worker in the form of lower wages, although part may be shifted to the consumer or may be borne by the employer. The cost imposed by the insurance is similar to the pay-roll tax imposed on the employer under the unemployment and old-age insurance programs. Insofar as shifting to the worker occurs, workmen's compensation amounts to compulsory accident insurance paid for by the worker.

The desirability of workmen's compensation acts is now rarely questioned. Insofar as producers and consumers of products involving high accident losses bear the costs of accident insurance, real social costs are being allocated to the industries insuring them. And even if the costs are shifted backward through lower wages, workers are usually better protected at less cost to themselves in this way than through any available form of private insurance. In either case there is reasonable assurance that injured workers' rights will be protected and that unforeseen accidents will not constitute complete disaster for low income wage earners. Compulsory employee accident insurance is sound social policy, even if workers themselves must pay the bill. Many enthusiasts for social re-

form, however, urge that the federal government should take steps to replace present arrangements by uniform nationwide protection.

Conclusion

The primary economic effect of the widespread social security program is to provide incomes for persons unable for some reason to earn incomes for themselves. Money incomes received under social security provisions give the recipients purchasing power and enable them to obtain a share of the national real income produced by employed labor and other productive agents.

Some social security expenditures contribute directly to increased future productive power—this is notably true of public health, child care, and rehabilitation expenditures. Establishment of employment exchanges and other steps to increase resource mobility act directly to increase current real income. On the other hand, the use of pay-roll taxes may directly lessen employment and reduce national income. And it is undeniably true that a large portion of social security expenditures in no visible way contribute to increased production, present or future. Such expenditures represent, therefore, simply a sharing of the national real income with nonworking persons by those who are still engaged in productive effort.

It is not necessarily an argument against social security measures that they provide security for some idle groups largely or even entirely at the expense of lower incomes for working groups. The whole social security program rests largely upon ethical and moral foundations; the premise is that it is the responsibility of society to care for those who for some reason are temporarily or permanently unable to care for themselves. It is argued also that redistribution of income in favor of the lower income groups should be undertaken for its own sake, quite aside from steps to increase the total real income to be divided. These objectives seem to be of growing importance in modern social thinking and policy making. In achieving this end of assisting particular needy groups and the lower income levels in general, the social security program has been reasonably successful, though in some cases only at the expense of reducing the total national income to be divided.

As a rapidly growing sector of the public economy, social security has exhibited serious internal conflicts. Set up primarily as a social reform program, it has been imperfectly coordinated and has often been in direct conflict with the public economy goal of business stabilization at relatively full employment. The institution of social security pay-roll taxes in 1937 was badly timed; indeed it helped to intensify and possibly precipitated the sharp business recession of that year. The repressive effect of the same taxes on employment during the "in between" years of 1938 and 1939 has been mentioned. Finally, with the inflationary pressure of war financing growing ever stronger in 1941 and 1942, many social security advocates fought all proposals to increase social security tax rates, however desirable such increases appeared as part of an anti-inflationary

fiscal program. Closer integration of technical social security practices with broad fiscal policies could surely be attained. Yet considerable rigidity in social security rates and payment practices may be justified on the grounds of keeping the system as far as possible out of the easy reach of political pressure groups. Were insurance premiums and payments put on an out-and-out pay-as-you-go basis, social security might easily degenerate into a political football. The conflicts of social policy cannot always be solved by the processes of economic logic.

Part XI

THE INTERNATIONAL ECONOMY

PREAMBLE

IN THE modern world it is hardly necessary to emphasize the close interdependence that exists between nations. History has demonstrated the economic advantages to be had from international trade and lending; it has also called attention to certain liabilities connected with heavy dependence on buyers and sellers abroad. To a considerable extent the real issues in international economics are those of political relations between nations, which have been and must be controlling over "purely economic" considerations when questions of war are involved. On the economic level, however, the case for extensive international trade and lending is remarkably clear, in spite of the consistent policies of trade repression followed by most countries.

Why do the layman and lawmaker generally distrust foreign trade and favor the tariff? Why have nations so consistently flown in the face of economic reason, if the case for international trade is so clear? There is probably nowhere else in economics a better opportunity for applying relatively simple economic analysis to popular fallacies predominant in public policy making. International economics, while fundamentally only an extension of the principles enunciated in the preceding Parts, constitutes an integral part of any body of economic doctrine that is to be useful in studying the real world of today.

CHAPTER 52

International Trade and Lending

Few indeed are the subjects in the field of economics on which the general public has been so long and so thoroughly misinformed as on the question of international trade and the tariff. Yet, strangely enough, there are few branches of economics where central principles are so simple and clear-cut, and where economists are so thoroughly in agreement as in this case. For the most part, the principles applicable to international trade are simply those of specialization and division of labor that underlie domestic economic activity. On the surface there would appear to be no reason to suppose that human welfare could be improved by obstructing the processes of specialization and exchange. Yet in fact, over long centuries governmental trade policy has consistently flown in the face of these principles. Such a phenomenon deserves careful attention.

The Basic Case for International Trade

Interregional and international trade

The advantages of interregional specialization and division of labor within nations are everywhere recognized. Iowa produces corn and hogs, Florida specializes in citrus fruits, and Alabama finds her circumstances well suited to cotton-raising. Few are the souls who would suggest that Iowa should raise her own oranges and cotton, Florida her own corn and pork. Differences in geography and climate, in the tastes and aptitudes of the people, and in the supplies of capital goods available, in different localities go far to determine in which lines of production each area can most advantageously specialize. Because these various resources and productive conditions are more or less fixed in the areas where they now exist, specialization occurs and goods are exchanged between areas. Were resources completely mobile, resources might move instead in accordance with consumer demands in different localities and little specialization and movement of goods occur, except insofar as prompted by advantages of large-scale production. But given widespread resource immobility, the most efficient economic distribution of resources will be obtained when, allowing for the given geographical distribution of human and material resources, each productive agent is used where its marginal productivity is highest. In this way the greatest quantity of wanted goods will be produced and residents of all participating areas will gain in the resulting exchange. In the simple example above, we might say that Iowa has an

"absolute advantage" in producing corn, and Florida an "absolute advantage" in producing oranges, since comparable amounts of productive resources will clearly produce more corn in Iowa than in Florida, more oranges in Florida than in Iowa. Under these circumstances it is obvious that both more corn and more oranges can be produced with interregional division of labor.

One major difference between interregional and international trade arises from the fact that many nations have different currencies and trade necessarily involves a currency-conversion in making payments. A second basic difference arises from the greater immobility of resources, especially human beings, in international trade. Interregional migration may be limited by institutional conditions but it is seldom flatly prohibited; by contrast, international migration is widely checked by national immigration barriers. But if these differences are waived, international trade constitutes merely an extension of the principles of interregional exchange. The mere existence of a national boundary line does not alter the basic principle that specialization and exchange are beneficial to the participants.

The law of comparative advantage ¹

If every region and every nation could produce every good equally cheaply, there would be no advantage in exchange between areas. However, different regions and different nations have very great differences in efficiency in producing different goods, differences that persist because of international immobility of resources. These differences arise largely out of five considerations:

1. Over the face of the earth climatic and geographical conditions vary widely. Brazil is admirably suited for raising coffee, Newfoundland is convenient to the fishing waters of the Grand Banks, and the rich soil of the lower Nile valley is ideal for cotton production. Minerals must be mined where the deposits are located. Thus Texas and Oklahoma are great oil-producing centers, while Chile has rich nitrate deposits. Such geographical and climatic differences alone would suffice to provide the basis for worldwide specialization, even though no other differences existed.

2. Just as soils and climates vary over the globe, so do human capacities. Some groups are physically large and strong, adept at physical labor. Others appear more successful at tasks requiring dexterity and manual skills. Still others excel in enterprise and organizational ability. These differences may be due to long-standing racial characteristics, or they may be due to the political, social, and economic environment in which individuals are raised and live. Whatever the reasons for such differences—and clearly differences in human capacities mirror a complex

¹ The exposition of this section follows that of R. F. Harrod, *International Economics* (Nisbet and Company, Ltd., London, 1939), Chapter II.

set of factors—their existence constitutes a major reason why international specialization and trade will be beneficial.

3. The accumulated supply of capital goods varies greatly from area to area and nation to nation. In some countries, centuries of accumulation have produced large supplies of fixed and mobile capital, in the form of railroads, buildings, machinery, and so forth. In others, where colonization and economic development have come more recently, stocks of capital goods are low. Moreover, vast differences exist between the kinds of capital accumulated in different nations. The United States is widely known for its vast, well-equipped industrial plants. Argentine capital, on the other hand, is largely in forms fitted to handling and processing beef, wheat, and other basic farm products.

4. The very uneven distribution of resources over the world means that the *proportions* between different types of resources vary widely from country to country. Thus Australia has vast plains but relatively few people and capital goods. Therefore, she finds it advantageous to specialize in products requiring a high proportion of natural resources to labor and capital goods. England, on the other hand, finds land exceedingly scarce relative to human beings and capital. Therefore, she is best fitted for manufacture and industry, even though her soil might be as good as Australia's for wheat-growing. In laying the basis for profitable exchange, such variations in proportions of the different type of resources are as important as differences in the resources themselves.

5. In addition to these more strictly "economic" considerations, great differences exist in the political and social climate in different countries. In countries of stable government and conservative traditions, vast industrial organizations requiring large long-period capital commitments are likely to grow up. By contrast, in backward, illiterate, badly governed areas, conditions are likely to be almost prohibitory against mass-production industry. A hustling, mechanical-minded nation such as the United States could hardly be expected to be satisfied again with a small-unit, predominantly rural economy, any more than the South Sea Islanders could be expected to be happy or efficient auto-makers. Perhaps political and social climate and traditions do not properly constitute a separate differing factor. Yet the importance of socio-political differences in determining national economic "aptitudes" cannot be overlooked, whether this factor be classified separately or merely reflected in the economic differences considered above.

Given these various differences between nations, it is clear that international trade will be advantageous. Similar differences underlie much of domestic trade, although domestically much more movement of resources and much less of finished goods is likely to occur. But to what extent each nation should specialize and how far international trade should go is not so obvious.

If the greatest possible advantages of foreign trade are to be obtained for all, each nation must devote itself to what it can do most cheaply.

In the oversimple case of Iowa and Florida producing corn and oranges, respectively, where the absolute advantage of each in its representative product is clear, both will benefit by complete specialization; to maximize real income Iowa should raise all the corn, Florida all the oranges. But the concept of absolute advantage is a slippery one when it comes to comparing costs between different nations having different monetary units, different proportions of factors of production, different labor standards, and different productive techniques. Most people would probably agree that as between coffee and factory machinery, Brazil has an absolute advantage in the former, the United States in the latter. Yet even in such an extreme case it is very difficult to be precise as to just why we are so sure that these absolute advantages exist. When less striking cases are considered, such as woolen manufacture in the United States and England, this difficulty becomes almost insurmountable. Monetary comparisons mean little, since different monetary units prevail in the two countries, and thus far no one has been able to find a satisfactory measure for comparing absolute costs in different countries. Moreover, even were such a measure available, the question would remain of the advantageousness of trade between nations where one has an absolute advantage in many or all products.

Fortunately, it is not necessary to say anything about the absolute advantage of different nations in determining how far specialization and trade can most profitably be carried. To illustrate this fact and the principle underlying the advantage from international trade, let us take a simple example, considering the position of one country, let us say, the United States, *vis-à-vis* the rest of the world with respect to the production of the two commodities, wheat and cloth.

Since comparative cheapness of producing different commodities is the issue at hand, it is necessary to have some unit for measuring cost of production within each nation. This may be dollars, labor-days, baskets of commodities given up, or whatnot—it does not matter what so long as some standard cost-unit is available. Having some unit of comparison, let us, so define units of wheat and cloth that the cost of producing one unit of wheat in the United States is equal to that of producing one unit of cloth. Say that each costs x cost-units to produce, as indicated in Table 52—1. This might mean, concretely, that with x cost one bushel of wheat or two yards of cloth can be produced. Then one bushel would be one unit of wheat, and two yards would be one unit of cloth.

Let the cost of producing the same unit of wheat abroad be y , again leaving undefined the exact nature of the cost-unit. This cost-unit bears no necessary relationship to the measure used in the United States; indeed, it cannot, for the reasons indicated above in discussing the concept of absolute cost. Costs in the United States may be measured in dollars, those abroad in man-hours; other standards are equally acceptable. Let us further suppose that the cost of producing our unit of cloth abroad is $2y$, assuming thereby that abroad the cost of producing cloth

is twice that of producing wheat. This situation is shown in Table 52—1.

Suppose that up until now no trade in wheat and cloth has occurred; each country produces its own supply of both. Will it pay to undertake further specialization and trade in the two commodities? The answer is yes. Productive resources are equally efficient in producing wheat and cloth in the United States; obtaining one more unit of wheat means giving up one unit of cloth, and *vice versa*. But abroad productive resources can be used only half as efficiently in producing cloth as in raising wheat; two additional units of wheat can be produced by giving up one of cloth. Therefore there will be an increase in total output if the United

TABLE 52—1

	<i>Cost of production in the United States</i>	<i>Cost of production abroad</i>
Unit of wheat	x	y
Unit of cloth	x	$2y$

States uses more of her resources to produce cloth while abroad resources are shifted to raising wheat. By transferring resources from cloth to wheat abroad, foreigners produce twice as many additional units of wheat as they give up of cloth; by transferring resources from wheat to cloth, the United States produces as many additional units of cloth as she gives up of wheat.

It might appear from the notation of Table 52—1 that the United States has an absolute advantage in production over foreigners, since abroad it takes $2y$ to produce a unit of cloth while here only $1x$ is required. Actually this has not been assumed at all. Since cost units x and y bear no necessary relation to each other, the table shows only

TABLE 25—2

	<i>Cost of production in the United States</i>	<i>Cost of production abroad</i>
Unit of wheat	x	$\frac{1}{2}y$
Unit of cloth	x	y

the comparative costs of producing wheat and cloth in the United States and the comparative costs of producing the two abroad. The situation might have been equally well indicated as in Table 52—2. Table 52—1 and Table 52—2 represent identical facts; exactly the same advantage from specialization and trade is indicated by both, as will easily be seen by tracing through again the resource shift discussed in the preceding paragraph.

The principle illustrated by these simple tables is that *gain from specialization and trade is possible if the cost ratios of producing two commodities are different in different countries*. This same principle would have applied in the above example had the cost ratios been 3:2 in the United States and 5:1 abroad, or any other set of differing ratios. It

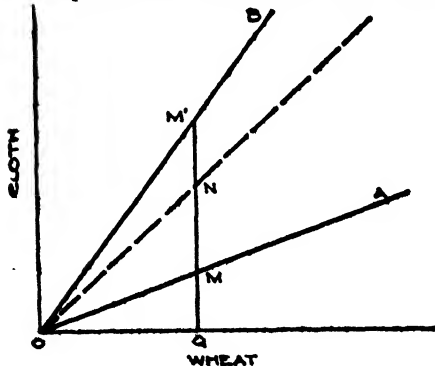
must be emphasized again that the principle and the tables say nothing about the relative costs of producing wheat here and abroad, nor of producing cloth here and abroad. The possibility of gain from trade does not depend on the comparative cheapness of producing each commodity as between different countries; it depends on the relative costs of producing wheat and cloth in one country in relation to the relative costs of producing them in other countries. Absolute costs are irrelevant.

In practical terms, for example, it would make no difference at all in the wheat-cloth case whether one day's work would produce one unit of wheat here and ten abroad, or *vice versa*, if it is assumed that the cost ratios are as indicated in the tables. Similarly, it would make no difference if one day's work here would produce five times as much wheat and ten times as much cloth as abroad. Gain from trade will occur whenever the cost-ratios differ. This principle, that gain will be maximized when each nation specializes in the lines where it has the greatest comparative advantage or the least comparative disadvantage, is commonly known as the "law of comparative advantage."

From the reasoning thus far it should be easy to tell how far it will be advantageous to carry specialization and trade. Gain from trade is possible until the cost ratios of producing the two commodities are the same in the United States and abroad. As production of cloth increases in the United States and that of wheat abroad, both sooner or later meet with increasing costs. In the United States the cost of producing cloth will rise relative to wheat, and abroad the cost of producing wheat will rise relative to cloth, until at some levels of output the cost ratios become identical here and abroad. Thereafter, there is obviously no advantage in further specialization and exchange, since no further increases in total output can be obtained thereby.²

As long as cost ratios are different here and abroad, either more wheat and cloth can be produced with the same resource expenditure, or the same amounts can be produced with less resource expenditure, permitting additional production of other commodities or increased leisure. But

² This principle of gain from trade can be shown in a simple diagram. Let one axis represent cloth and the other wheat. Any point on *OA* shows the relative cost of producing wheat and cloth in country *A*, while any point on *OB* shows the cost ratio in country *B*. Thus for some output of wheat, *OQ*, in *A* an identical cost is required to produce *QM* cloth. In *B*, the cost of *OQ* wheat will produce *QM'* cloth. Therefore, clearly gain will accrue if *A* specializes in wheat and *B* in cloth.



As this specialization proceeds, wheat will be subject to increasing costs in *A* and cloth to increasing costs in *B*. Thus lines *OA* and *OB* will move toward each other. When they become identical, let us say along the dotted intermediate line, the cost ratio will be *OQ:QN* in both countries, and further gain from trade will be impossible.

the extent of the gains from trade and how they will be divided between the countries involved are more difficult questions. In general, we may say that the greater the difference in cost ratios the greater the gain from trade that is possible by increasing specialization until the ratios become identical. The exact division of gains among the countries concerned depends largely upon the cost functions and demand conditions for both products in the various nations, and upon the size of each country relative to the others. But the problem of measuring these gains is so complex that it does not, for our elementary purposes, repay detailed study.

When the law of comparative costs is generalized to the real world with its many countries and thousands of products, no new principles are introduced but the picture becomes tremendously more complex. Gain will still be maximized if each country specializes in producing those goods and services in which its comparative advantage is greatest or its comparative disadvantage least, carrying this specialization up to the point where the cost ratios involved are equal to those of other countries producing the same products. This means, of course, that in any given country production of many products will never take place, because the country's comparative disadvantage in their production is so great. It means, likewise, that each nation will almost certainly find it advantageous to produce a considerable variety of products; even though its comparative advantage may be unusually great for some particular product, the advantage will seldom be large enough to exclude production of all other goods, especially since expanding output sooner or later means rising unit costs. The introduction of transportation costs into the picture further strengthens the likelihood that each commodity will be produced in several countries and each country will produce a variety of products. Trade is profitable only when the gain is sufficient to overcome transport charges. Therefore the existence of shipping charges, insofar as they are greater internationally than domestically, discourages foreign relative to domestic trade.

The law of comparative costs, free trade, and the price system

With free trade and free, competitive movement of prices in all countries, the maximization of gain indicated by the law of comparative costs would tend to be realized as entrepreneurs and individual resource owners in every country sought to maximize their incomes. In each country, under such circumstances, the greatest profits could be obtained by increasing production of those commodities most desired by consumers at home and abroad. If the United States' comparative advantage is great in producing automobiles, this fact would be reflected in high returns to resources in the automobile industry. If our comparative advantage is low in producing spices, American spice producers would be unable to pay wages high enough to bid resources away from the more efficient automobile industry, since the world price for spices is set by relatively efficient East Indian production. Thus under a free price system and free trade, the resources of each nation would tend to be

drawn through the search for profits into its relatively most efficient industries, and exchange would tend to occur in the manner indicated in the preceding section. The details of this process, however, must wait upon the discussion of differing national monetary systems in the next chapter.

Insofar as barriers are established to the free international movement of goods or the free movement of resources within any nation, the process of specialization and trade is impeded and the gains from trade lessened. The case for free trade internationally is the same as the case for free trade domestically—that, in the absence of restrictions, resources will presumably be most advantageously allocated by the price system in accordance with consumers' demands and resource owners' preferences as to jobs. In foreign, as in domestic, trade, a variety of circumstances may arise under which one or a group of nations may gain by restricting trade. Some of these situations call for special attention in subsequent chapters. But on balance they bulk small against the basic advantages of free exchange.

The principle of comparative costs specifies the conditions under which the optimum allocation of resources within each nation may be obtained. It does not deal with the problems of resource unemployment. This does not mean that the principle of comparative costs rests necessarily on the assumption of full employment, but rather that it takes the level of employment as a given, determined largely by other considerations. The relation of foreign trade to domestic business fluctuations must be considered subsequently, but its exclusion here need not disturb us greatly.

The Basic Case for International Lending

Society benefits when individuals and institutions save and invest, because useful capital goods and durable consumers' goods are thereby accumulated. The saver gains individually from the return on his investment, and society gains from the increased efficiency of roundabout production permitted by additional capital goods or directly from having more consumers' goods. And both the saver and society gain most when savings are invested where their productivity is highest.

In the domestic economy we have for the most part trusted the price system to allocate savings to the most desirable investments. Savers (directly or through such institutions as banks and insurance companies) invest their funds where the combination of safety, liquidity, and rate of return seems most attractive. Since the rate of return tends to be highest in those investments that directly or indirectly fulfill the strongest consumers' demands, savings will thereby be drawn into those industries where consumers most want output increased. Thus, although the process works imperfectly, consumer choices direct the allocation of savings among investments, with the rate of return available in different investment opportunities the basic allocation guide.

Internationally, as well as domestically, society is generally best off

if savings are allocated to those investment opportunities where their marginal productivity is greatest. International loans are "better" than domestic loans when the anticipated rate of return on them (including risk allowance) is greater. For decades both England and the New World gained from heavy loans by Britishers to individuals and businesses in the rapidly developing Western Hemisphere. British lenders gained by receiving good returns on their investments; borrowers gained by obtaining additional capital to combine with the plentiful natural resources of the New World. This same process of lending by older, wealthier nations in undeveloped areas has gone on over several centuries, although growing barriers have recently stifled such lending. Internationally, as well as domestically, lending may be unwise if the loan is unsafe, or undesirably illiquid, but internationally as well as domestically both lender and society tend to benefit most from having savings converted into those investments that promise the greatest return. Here again the existence of national political boundaries does not invalidate the basic economic principles of gain from exchange and lending.

International Community and Conflict of Interests

In general the conclusion that every country gains from free trade and from the increased efficiency and welfare of other countries holds good. However, there are certain cases that prove exceptions to this proposition. Some of these, especially those that may arise as temporary phenomena, will be noted in connection with the later discussion of tariffs. Some of them are, however, of more fundamental importance and need to be specifically considered here.

Perhaps the most important conflict between the world interest and the purely national interest arises in connection with the problem of personal migration. If the "economic welfare" of the world as a whole is the ultimate aim, then international migration probably should occur whenever the real wages obtainable in one country are higher than in another. This is merely an application of the proposition that mobility of resources conduces to efficient use in accordance with consumers' demands. But for the workers obtaining high wages in some particular country, let us say the United States, an immigration of workers from low-wage countries might prove a definite and possibly major financial blow. Such an influx, if it is assumed that the immigrants were fairly readily substitutable for American workers, would directly lower the wages and incomes of American workers. The decline in wages in the United States would actually not go as low as present levels in other countries, because of the superior complementary American resources and technology and because a considerable part of the high American wages rest on superior training and abilities of American workers, so that foreign labor would ordinarily be a far from adequate substitute. But this conflict of national and international interest should not be glossed over by loose application of the law of comparative advantage. Given the international distribution of resources, the law of comparative advan-

tage applies; but the law does not say that each nation will gain from an international shift of resources, even when the shift raises the average plane of living of the world taken as a whole.

Another possible conflict of national and international interests arises in connection with lending for the development of "new" countries. If the borrowing country uses the funds to develop industries directly competitive with those of the lending nation, the lender may eventually find itself worse off because of the loans. While this result has often occurred, frequently the types of goods produced by the developing country are so different from those of the mature country that the cases of direct competition are overshadowed by those of a less competitive nature. And the more products the new country produces for international exchange, the greater its purchasing power to buy the products of the lending country.

This same set of considerations apply to any case of differential rates of international development. The economic progress of one country is ordinarily a benefit to others, through increasing the world supply of goods and the developing country's international purchasing power. However, if the expanding nation's products compete directly with those of a less rapidly developing or declining country, the economic gain of the former forces down rewards to productive agents in the latter if the latter is to hold its share of the world market.

Lastly, conflict may arise in cases where one country has adopted minimum-wage laws, maximum-hour provisions, or other social reform measures that raise labor costs per unit of output, or where trade-union pressures have had this same effect. In such cases, if there is free trade, low standard countries will be able to undersell nations with superior labor standards, thereby forcing down wages or eliminating the special provisions in the more advanced countries. Countries with advanced labor standards, such as England and Sweden, are therefore necessarily interested in promoting higher standards in other countries, to protect their own interests. It should be noted that this problem of differential standards is not peculiar to the international situation; it has, for example, arisen widely within the United States, as between sweat-shop and unionized garment workers.

Political Aspects of International Trade

Economic factors have been of great importance in determining the course of international political relations, though probably many persons overestimate the importance of purely "economic" considerations. It would be desirable to discuss at length the complex interrelations between international economic and political factors, but unfortunately this cannot be undertaken here. It must suffice to say that since here only the "economics" of international trade and finance is considered, only part of the picture is being recognized; and that noneconomic considerations may well at times play the dominant role in dictating international commercial and financial policy. In general, however, it is probably safe

to say that economic and political considerations point in the same direction, with major exception of the desire for national self-sufficiency in expectation of war. Free international trade and finance are probably the soundest possible bases for lasting peace; and lasting peace may well be unobtainable without such free economic intercourse. Conversely, international peace and goodwill are the necessary basis for free international economic intercourse. But in a politically unstable world, sacrifice of the economic advantages of free trade may seem a small price to pay for national self-sufficiency, even though economic nationalism aggravates political antagonisms.

The Balance of Payments

International economic relations consist not only of commodity exports and imports but also of exports and imports of services (such as shipping and banking services) and of capital. Table 52—3 shows the complete "balance of payments" of the United States in 1939, a somewhat abnormal year in view of the beginning of the war in its late months, but on the whole typical of the last decade. The left-hand side (credits) shows items that involve payments from other countries to the United States.

TABLE 52—3 *

UNITED STATES BALANCE OF INTERNATIONAL PAYMENTS, 1939 (000,000's omitted)

CREDITS		DEBITS	
<i>Trade and service items:</i>			
U. S. merchandise exports . . .	\$3,117	U. S. merchandise imports	\$2,318
Shipping and freight services sold	125	Shipping and freight services bought	249
Tourist expenditures by foreigners here	159	Tourist expenditures by Americans abroad	469
Immigrant and institutional remittances	45	Immigrant and institutional remittances	187
Interest and dividend payments	531	Interest and dividend payments	211
<i>Capital movements:</i>			
Long-term capital movements	1,624	Long-term capital movements	1,510
Short-term capital movements (net)	1,116		
<i>Gold and silver movements:</i>			
		Gold and silver movements to U. S. (net)	3,111
<i>Miscellaneous:</i>			
Miscellaneous and uncounted items due to U. S. (net)	1,338		
	<u>\$8,055</u>		<u>\$8,055</u>

* Data from *The Balance of International Payments of the United States in 1939*, Bureau of Foreign and Domestic Commerce (Washington, 1940), p. 6.

The right-hand side (debits) shows items that involve payments from the United States to other countries. Since the payments due from other countries exceed those due from us to other countries, the difference was paid by them in gold and is shown on the right-hand side, under "net gold movements." Very few of these items involve transactions directly between governments; most of them represent transactions between individuals and businesses engaged in foreign trade and investment.

It is necessary at the outset to clarify the distinction between the "balance of payments" and the "balance of trade," a matter on which much popular confusion exists. The balance of payments includes *all* payments between the countries concerned; the balance of trade includes only the trade in goods and services, omitting "financial" transactions such as foreign investments, interest payments, and immigrant remittances. Most popular discussion runs in terms of the balance of trade, and often the reasoning is confused because of failure to consider the entire balance of payments, including "financial" transactions. For example, the heavy gold inflow into this country during the 1930's was due in part to a "favorable balance of trade" but probably more to purely "financial" transactions involving the flow of funds to this country for long- and short-term investment and sometimes simply for safekeeping. The distinction between the balance of payments and the balance of trade must be kept clear if discussion of international trade policies is to be cogent.

Table 52—3 shows that during 1939 exports of merchandise from the United States (\$3,117,000,000) exceeded imports of merchandise (\$2,318,000,000) by some \$859,000,000. On the other hand, American tourists spent a good deal more abroad than foreigners spent here. Also foreigners sold us more shipping and freight services than we sold them. And immigrants, other individuals, and institutions in this country sent back to friends, relatives, and needy persons abroad a sum considerably larger than similar payments to this country. These items of tourist expenditures, services, and remittances cut down by about \$520,000,000 the net amount owed by foreigners to people in the United States, although there remained a large net balance due to the United States. When a country exports more goods and services than it imports, receiving gold to settle the balance, it is often said to have a "favorable balance of trade."³ Actually, as will become evident, there is nothing necessarily favorable about a "favorable balance of trade," but this term was applied by the "mercantilists" (who looked upon the inflow of gold as of primary importance) in the seventeenth and eighteenth centuries, and the usage of the words has carried over to the present day.

This net balance due the United States in 1939 was increased by financial transactions. The balance sheet shows that interest and divi-

³ Remittances, being merely financial transactions, are not part of the balance of trade.

dends paid from abroad to United States investors greatly exceeded interest and dividends paid to foreigners on their investments here. The United States is internationally a creditor country—its residents have more investments abroad than other countries' residents have here. This position is the result of the great foreign loans made by the United States government and by American citizens during the First World War and all through the 1920's, especially to Europe. Net interest payments increased the balance due this country by \$320,000,000 in 1939.

But long- and short-term "capital movements" account for the largest single part of the net balance due the United States. On long-term investments (such as bonds and investments in buildings and businesses), funds sent here by foreigners exceeded by some \$114,000,000 new American investments abroad. And the flow of short-term capital (such as bank deposits and investments in very liquid securities) into the United States exceeded the corresponding outflow by about \$1,116,000,000. In large part, this heavy capital inflow, which characterized the 1930's, resulted from unsettled political conditions and then war in Europe. Most of the funds came here because their owners considered the United States a safer place than war-torn Europe; others came merely in search of profitable investment opportunities.

The huge inflow of gold into the United States in 1939 thus came primarily as the result of three factors: First, on balance foreigners owed people in the United States about \$323,000,000, because they bought more goods and services from us than we bought from them. Second, on investments already made, interest and dividend payments to the United States exceeded outgoing payments by another \$320,000,000. Third, primarily owing to unsettled political and economic conditions in the rest of the world, foreigners sent large amounts of short- and long-term

TABLE 52—4

UNITED STATES BALANCE OF PAYMENTS

January 31, 1934 to January 1, 1941

(All figures net)

Net credit on merchandise exports	\$ 4,365,000,000	
Less: Net debits on service items	954,000,000	
Net credit on merchandise and service items		\$ 3,411,000,000
Net short-term capital inflow	4,938,000,000	
Net long-term capital inflow	2,176,000,000	
Net capital inflow		7,114,000,000
Unidentified items—net credit		4,643,000,000
		<u>\$15,168,000,000</u>
Net gold and silver inflow:		
Net gold inflow	\$14,142,000,000	
Net silver inflow	1,026,000,000	
		<u>\$15,168,000,000</u>

capital here for investment and safekeeping. Such capital movements accounted for another \$1,200,000,000.

That the year 1939 is fairly typical of the last decade following the 1933 depression low is shown by Table 52—4, which gives the net credit and debit figures for the United States over the period January 31, 1934, to January 1, 1941. This simplified long-period table shows clearly the two basic causes of the gold (and silver) inflow over the period. It is important, however, to keep this period in proper historical perspective. Never before had the world seen such huge gold flows; never before had there been such huge panic-stricken "flights" of short-term capital. By the end of the period shown in the table, the rest of the Western World

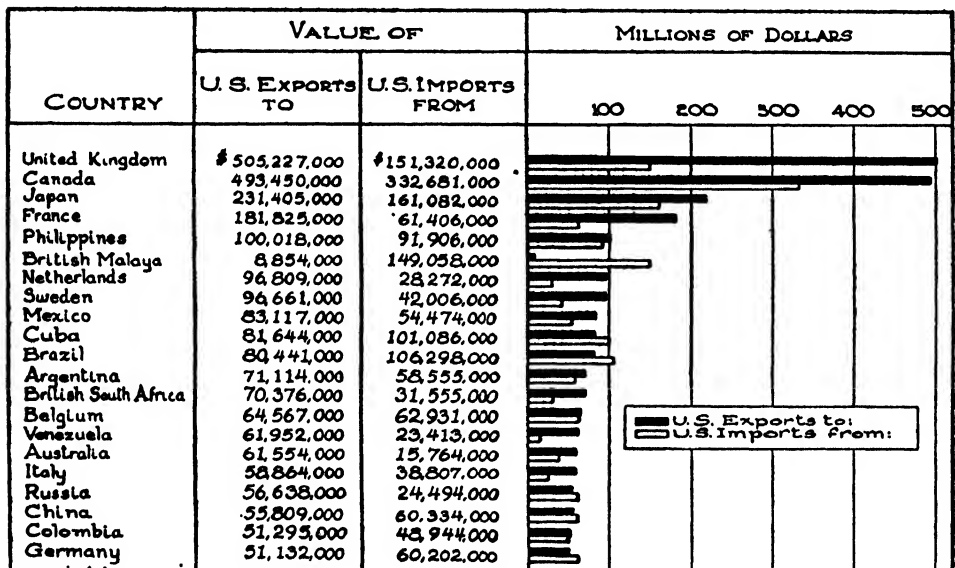


Fig. 52—1. International trade of the United States in 1939.

was largely drained of its monetary gold except for small reserves, and with the development of worldwide war, gold flows were restricted to small scope. What they may be in the postwar period will depend on the development of the entire international balance of payments.

The volume of international trade between the United States and the principal countries with which we traded in 1939 is shown in Figure 52—1. This table shows only commodity exports and imports—it does not include services or capital and interest payments. Except for British Malaya, countries are listed in order of their importance as buyers of United States products. Figure 52—2 shows the actual commodities that were our most important exports in 1939. And Figure 52—3 shows the commodities that were our most important imports in the same year. It should be noted that it makes some difference in the rank of commodities as to how they are grouped. For example, if all machinery were grouped under one head, it would stand first on the export list.

Raw materials are predominant among the imports, while finished and semifinished products bulk large among the exports. Our greatest comparative advantage seems to have shifted away from farm products (where it was through the early period of the United States) toward manufactured products, though some raw materials (especially cotton) remain very important exports.

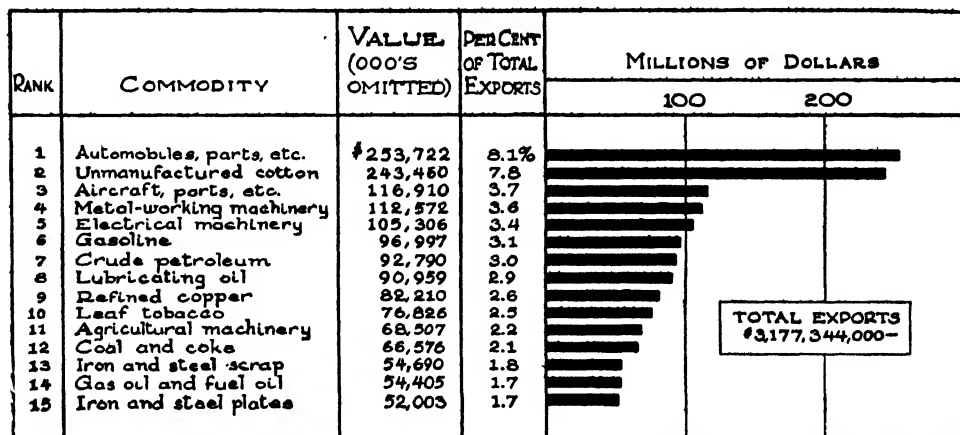


Fig. 52—2. Chief exports of the United States in 1939.

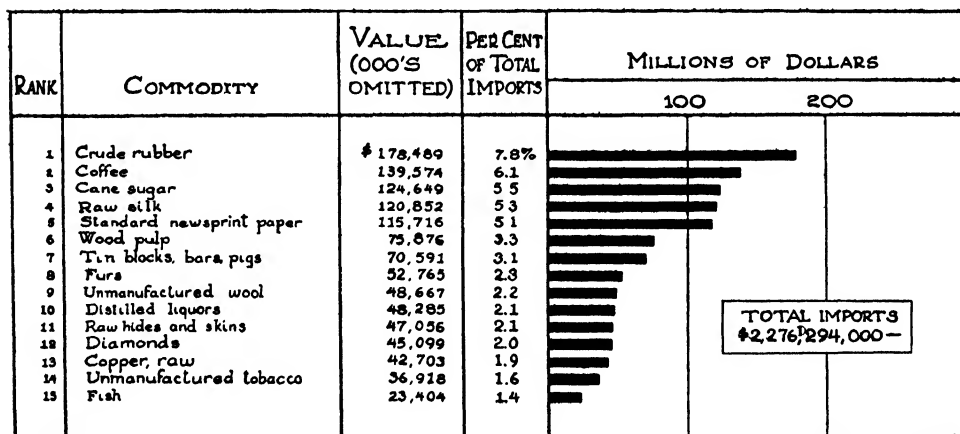


Fig. 52—3. Chief imports of the United States in 1939.

The 1939 net capital inflow of about \$1,200,000,000 (not shown on a chart) came predominantly from Europe, which accounted for over \$700,000,000 of the total. Within Europe, Switzerland, the Netherlands, and France were the most important single sources, while England, facing huge war outlays, was the only country showing a new withdrawal of capital. Of the remaining capital inflow, about \$135,000,000 came from Asia; \$72,000,000 from Canada; \$94,000,000 from Latin America; and the rest from miscellaneous sources.

Of course the make-up of the balance of payments varies widely from year to year and over longer periods. The dollar volume of payments has also varied sharply, although the general tendency has steadily been upward during the last two centuries. Two factors have been dominant in causing repeated sharp fluctuations in the volume of international transactions: (1) unsettled world economic conditions, especially during severe international depressions and booms; and (2) unsettled world political conditions, especially in connection with wars. Thus current figures would show much larger exports and imports than in 1939, but the earlier figures give a picture less distorted by war.

How important is international trade to the United States? During the prosperous 1920's, roughly 10 per cent of all goods produced in the United States were exported. During the depression this figure fell off to somewhere around 5 per cent, partly as a result of bad business conditions and partly as a result of growing trade barriers here and in other countries. By the beginning of the present war exports had again increased until they were approaching the 1920's percentage, and since 1939 a sharply increasing volume of exports has been sent abroad, to aid our allies, to supply United States forces abroad, and to exchange for materials needed for American war production. With increasing reliance on "Lend-Lease" assistance, the traditional pattern of sales of goods abroad has given way more and more to what are in effect gifts of war materials by the United States government to our allies.

While the peacetime percentages shown in the preceding paragraph may seem to indicate that international trade has been relatively unimportant, one must not generalize too hastily. Even 5 per cent to 10 per cent is an important part of total production, and numerous particular commodities included in international trade are peculiarly vital to American welfare. As high as 70 per cent of the total cotton crop, for example, has been exported in certain years, with the annual average over 50 per cent; foreign markets largely determine the welfare of cotton producers. The failure of foreign markets is thus a basic factor explaining the plight of the cotton South over the past decade. Although cotton is perhaps the most extreme case of dependence on foreign buyers, many other agricultural and industrial products rely heavily on export markets. Similarly, as has been demonstrated dramatically by war shortages, many domestic industries have been heavily dependent on foreign-produced raw materials (for example, rubber), and it is virtually impossible to obtain some consumption goods (for example, coffee) except from abroad. In terms of the maladjustments created by loss of foreign markets and suppliers, foreign trade is even more important than indicated by the total percentage figures for foreign and domestic trade.

CHAPTER 53

Financing International Transactions

ALTHOUGH they are essentially similar, one crucial difference between international and domestic trade is this: in domestic trade one common currency is used, but in international trade the participants in any transaction are likely to use different currencies. If a British woolen manufacturer sells cloth to an American importer, the American would have United States dollars to pay but the Britisher would want English pounds. This problem is common to most international payments—there must be a conversion of currencies. Because international trade and capital movements are of great importance, and because the problem of different currencies is peculiar to them, it is useful to analyze briefly the mechanism by which international payments are carried out under various conditions. Recently this subject has attained added importance since manipulation of the ratio of exchange between different currencies has been widely used in controlling international trade and even in regulating domestic business conditions.

Foreign Exchange

The financing problem in international trade arises primarily from the fact that the currency the buyer has is likely not to be acceptable to the seller. The British woolen manufacturer wants British pounds, not American dollars, because the things for which he wants to spend the money in England must be paid for in British money. Only if it is possible for the American's dollars to be converted into English pounds can the business transaction between them be completed.

Because there is this continuous demand for the conversion of various currencies in carrying on all sorts of international transactions, certain institutions have made this their business, charging a commission on each conversion. In the United States these institutions are primarily big New York banks; in other countries large firms have grown up with this exclusive function. Other banks or financial institutions, receiving demands from customers for conversion of currencies, simply pass on the transaction to one of these regular dealers.

Let us trace through a very simple transaction of this sort.¹ Let us say

¹ This example is not an exactly accurate statement of the manner in which such a transaction would probably be executed. However, the actual procedure, while often much more complicated, follows the same general principle.

an American importer, Mr. X, has agreed to pay a Britisher, Mr. Y, £1,000 in British money for some woolen goods. He therefore goes to, let us say, the Chase National Bank in New York, and announces that he wants to buy 1,000 British pounds. The Chase National being a regular dealer in such conversions, agrees to sell him the £1,000 at the rate of, say, \$5 for £1. Mr. X then pays \$5,000 plus a commission to the bank and receives a bank instrument, quite similar to a regular bank check or draft, giving the holder the right to £1,000. This document, called "foreign exchange," he sends to Y, who can take it to his bank in London and receive for it £1,000. The London bank then sends the foreign exchange note to the branch maintained by the Chase National in London, and receives back the £1,000 paid out to the British woolen manufacturer. So the transaction is completed—X has paid for the woollens with American money; Y has received English pounds in payment; the Chase National, acting as middleman, has \$5,000 plus a commission in New York, and has paid out £1,000 from its London branch. The existence of many such foreign exchange dealers in different countries provides a constant conversion market for the world's major currencies.

The rate at which British pounds can be bought with American dollars is called the "rate of exchange," or the "exchange rate." When X buys pounds with dollars, he is purchasing foreign exchange. Foreign exchange is merely a claim on some foreign currency, and the rate of exchange is the number of units of one currency it takes to purchase one unit of another currency. Put in other words, the exchange rate is simply the price of one currency in terms of some other currency—for example, if the dollar-pound exchange rate is 5 to 1, the price of one pound is five dollars. If the dollar franc exchange rate is 1 to 33, then the price of one franc is about three cents.

Since the exchange rate is nothing but the price of one currency in terms of another, an increased demand for any currency for making international payments will tend to raise its price in terms of other currencies, and a decreased demand will lower its price in terms of other currencies. For example, suppose American importers need large numbers of pounds to pay British manufacturers. When they try to buy these pounds with dollars, this increased demand will force up the price of pounds—the dollar-pound exchange rate will move in favor of the pound, say to 6 to 1, and the price of a pound would become six dollars. Conversely, when Britishers want to send capital here for investment or safekeeping, they buy dollars to invest or deposit here, thereby bidding up the price of dollars, so under those circumstances the exchange rate might move to 4 to 1. One pound would then buy only four dollars, or, what is the same thing, it would take only four dollars to buy one pound.

What the exchange rate will be at any given time, then, depends on the supply of, and demand for, the currencies in question. The conversion demand for any currency (let us say, United States dollars in England) comes primarily from (1) British people and firms who have bills to pay in the United States, and (2) British people and firms who

desire to send capital to this country. Thus the demand for dollars in England would come primarily from importers there who had bought goods in America, businesses there that owed interest and dividends on American investments in Britain, and British people and businesses who wanted to send capital to America for investment and safekeeping. Across the ocean, in New York, the American demand for pounds arises from the same sort of circumstances.

To illustrate the exchange process somewhat more fully, let us expand the previous example of Mr. X's payment of \$5,000 (£1,000) through the Chase National to Y in London. This was a demand for £1,000 in New York. Suppose now also that a British importer, A, owes \$5,000 to an American cotton wholesaler, B. Mr. A goes to the Chase National branch in London to purchase \$5,000 of foreign exchange to pay his debt. At the 5 to 1 exchange rate he would pay £1,000 for the dollars, his purchase constituting a demand for dollars in London. This \$5,000 of "dollar exchange" A would send to B in America, who would get \$5,000 for it from the Chase National in New York, thereby completing this second transaction. Thus the Chase National in New York would have received \$5,000 on the first transaction and paid out \$5,000 on the second. In London, its branch would have paid out £1,000 on the first transaction and received £1,000 on the second. In this simple illustration, the whole international exchange of payments is carried out without any international shipment of money (gold); both debts are settled, and the Chase National is ahead by the commission received on each transaction. In effect the debts have simply been canceled off against each other.

Of course, Mr. A might not have gone to the Chase National branch in London, but instead to some other foreign exchange dealer. This would have caused no difficulty. In both London and New York there are regular foreign exchange markets. If the Chase National in New York has large demands for pounds, but a small supply of pounds in England because few people have happened to buy dollars at its London branch, it could simply go into the New York exchange market and buy pounds from other dealers who happened to have larger supplies on hand in London. As long as the total demand for pounds in New York equals the total demand for dollars in London, all transactions can be settled without any change in the exchange rate or any international shipment of gold. The mere fact that any one particular dealer fails to have just the same demand for dollars in London as for pounds in New York is immaterial, since the total supply meets the total demand in the regular foreign exchange markets of each country.

Exchange Rates

Exchange rates with "free" currencies

The statement that exchange rates depend on the relative demand for, and supply of, the currencies in question, is applicable to all cases, except

those where exchange rates are flatly fixed by government decree, as at present in many countries under war conditions. Even where governments have fixed exchange rates, under-cover, unofficial dealing in foreign currencies has usually developed in "black markets," where rates are fixed by the relative supply and demand forces. For the present this case of government fixing of rates is excluded and attention is centered on cases where there is no such government control. An understanding of the determination of exchange rates is essential to an understanding of the processes of international trade and the exchange control devices recently widely used to control international trade and stimulate domestic business conditions.

Suppose now that the two currencies under discussion are completely free to move in response to supply and demand forces. In this simplest case, an increased demand for dollars in England relative to the supply of dollars will result in a rise in the price of dollars, a fall in the price of pounds. In such a case the exchange rate might change to, say, 4 to 1, so that one pound would now buy only four dollars, instead of five. Such a shift in exchange rates in favor of the dollar and against the pound might be expected whenever Britishers were sending more funds to the United States (for whatever reason) than Americans were sending to England. For example, since the United States typically exports more to England than we import from her, other things equal the exchange rate would move in favor of the dollar as noted above, if exchange rates were "free." Similarly, the great international capital flow to the United States during the 1930's tended to increase the price of dollars relative to other currencies. With completely "free" exchanges, exchange rates at any time are determined by relative demands for, and supplies of, the currencies in question.

Exchange rates under the gold standard

During a long period before the First World War, and again after the war, most important countries were on the gold standard. During these periods no direct control was exercised over exchange rates by governments, but owing to the nature of the gold standard exchange fluctuations were held within narrow bounds. Under the gold standard, the monetary unit "contained" (was convertible into) a certain number of grains of gold. For example, the prewar dollar was 23.22 grains and the prewar pound 113 grains of fine gold. Thus the pound had 4.86 times as much gold as the dollar, and it was this relation that established a par of exchange between the currencies.

Under these circumstances, suppose there was an increased demand for dollars, which sent the price of dollars up so that one pound would buy less than \$4.86 (say the exchange rate moved to 4.80 to 1). Then instead of buying dollars and getting only \$4.80 for £1, Britishers could simply convert their pounds into gold, send the gold to America and there get \$4.86 for the 113 grains of fine gold in each pound. Obviously, therefore, the exchange rate could not vary far from 4.86 to 1 or gold

would be shipped instead of foreign exchange being used at all. Actually, the cost of shipping the precious metal and the loss of interest on it while in transit made it unprofitable to ship gold unless the exchange rate varied more than 3 cents either way from the 4.86 to 1 ratio. Hence exchange rates could fluctuate within these "gold points" of 4.89 and 4.83, but no farther. Under the international gold standard, the relative gold contents of the various currencies governed their exchange rates, except for minor fluctuations within the gold points.

This stability of exchange rates under the international gold standard was a great boon to international traders, for they were always certain within a narrow range just what would be the value at which they could buy foreign currencies to make payments abroad and at which they could convert foreign currencies received in payment into domestic currency. Whatever the disadvantages of the gold standard in tying monetary systems and price levels rigidly to gold so that monetary policy against business cycles was hamstrung, the policy rendered a significant service in providing a stable currency-conversion basis for international trade and lending. However, stable exchanges may be obtained without having a gold standard. Moreover, if only one country is on the gold standard, this does not assure stable exchange rates because there is not conversion of other currencies into gold at any fixed rate. More will need to be said about these points presently.

"Controlled" exchange rates

Today, since the traditional form of the international gold standard no longer exists, there are no gold points to set limits to exchange rate fluctuations. On the other hand, there are virtually no cases where exchange rates are left "free" to move up and down as sharply as dictated by varying private demands for, and supplies of, exchange. All important cases fall into the intervening classification of exchange rates unfixed by gold points but controlled at least to some extent by government action of various sorts. The ways in which these government controls work vary greatly, and only a few of the most direct and common methods are considered here. During the war such controls have become increasingly stringent.

1. **"Stabilization funds."** Perhaps the most common, and the least drastic, method of governmental control of exchange rates has been through government intervention in the exchange market as a buyer and seller to mitigate undesired fluctuations. For example, if the United States government sees that the dollar is depreciating undesirably in exchange value, it may simply enter the foreign exchange market and buy dollars to check the fall of the dollar in terms of foreign currencies. Conversely, if the dollar is appreciating and the government wishes to maintain a stable exchange rate, it can sell dollars to hold down the price of dollars. This sort of government intervention is usually carried on through a special fund set up for the purpose, generally called an

"equalization" or "stabilization" fund. In the United States, such a fund was set up in 1934 with \$2,000,000,000 of the "profits" arising from the devaluation of the dollar, and it has been operating ever since, though it intervenes in the market only seldom. On the other hand, the British Equalization Fund has been much more active, as have been those of other countries. Obviously, the limits to the control any equalization fund has over exchange rates depend on the resources it has at its disposal to buy and sell exchange.

Such equalization funds may act to hold exchange rates stable, to raise them, or to lower them, depending on the result desired in any particular case. Quite often, exchange rates are temporarily "pegged" at some level—that is, the equalization fund enters the market whenever necessary to keep the rate at just the desired level. For example, from 1936 until the beginning of the war, the United States, France, and England, plus some smaller countries, had an agreement (called the "Tripartite Monetary Agreement") to keep exchange rates between their currencies pegged at about the existing levels. Whenever rates would begin to move upward or downward much, the various equalization funds would buy or sell exchange to keep rates at existing levels. Such a stabilization agreement provides stable exchanges to facilitate international economic intercourse, much as did the gold standard.

2. "Blocked exchange" and "exchange rationing." One of the earliest steps in German exchange control was the introduction of "blocked exchange." Persons outside Germany receiving funds in Germany, let us say as interest on bonds, could not convert those funds into foreign exchange for withdrawal. Instead the funds were "blocked" in Germany, where they had either to be spent or simply held in hope of future withdrawal. This special restriction on foreigners led rapidly to similar control of the use of exchange by German nationals, culminating in complete control of all purchases and sales of foreign exchange by government permits or licenses. Under this (the present) arrangement, no one can obtain foreign exchange to make purchases abroad except through the government. In the same way, all exports are licensed by the government, and foreign exchange received from foreign sales must be turned over to the government for use as public authorities see fit. This device has been used primarily to conserve the supply of foreign exchange (dollars, kronen, and so on) for use by the German government in buying arms and other "essential" materials abroad. The process by which available foreign exchange is allocated to persons wishing to make payments abroad under such an arrangement is called "exchange rationing," since the government rations the available exchange to those persons who it thinks will use the foreign funds most advantageously.

Where exchange rates are closely controlled and held at artificial levels by government action, as is now generally true outside the Western Hemisphere, almost everywhere "unofficial" under-cover markets for foreign exchange have developed. Thus there may be simultaneously the "official" prescribed rate and the "unofficial" rate, which moves more

freely in conformity with forces of supply and demand. In Germany dealing in such an "unofficial" manner in foreign exchange is a major crime; in England such dealings have been quite open. Such "unofficial" rates and the extent of their divergence from "official" rates may be of considerable importance in influencing international trade, but the complications involved are so great as to prevent further discussion here.

3. Exchange clearing. Another sort of exchange control that became important in Europe during the last decade is "exchange clearing." Under this arrangement the regular foreign exchange market is eliminated. The countries making an exchange clearing agreement agree that each will import from the other some specified amount during the time period covered. Let us say that the countries are Switzerland and France, and each agrees to import from the other one hundred million francs worth of goods in a year. Then all Frenchmen buying from Switzerland turn over their payments to the French government instead of sending exchange to Swiss sellers; and all Swiss buying from France turn over their payments to the Swiss government. Next the governments "clear" these debts, just as under a bank-check clearing system, and cancel their equal debts to each other. Then the French government uses the hundred million francs received from French importers to pay the French exporters who did not receive payment from Swiss buyers; the Swiss government does likewise within Switzerland. Thereby the debt between countries is canceled with only transfer payments within each nation. Exchange clearing systems may be set up between several countries as a group rather than just between two, but most agreements have been bi-lateral. Although sometimes they provide for increasing trade between participants, such exchange clearing agreements are a definite step backward toward a barter system of international trade.

In later chapters the significance of these various types of exchange control as devices for regulating international trade and for affecting domestic business conditions and employment will be examined. Some countries have gone so far as to eliminate completely the use of foreign exchange or any sort of money in some of their international dealings, turning instead to direct barter. Germany, for example, before and during the present war, has concluded many agreements with other countries to barter German manufactured goods for desired raw materials and foodstuffs. This return to barter is the extreme case of control, with far-reaching repercussions on the domestic economies involved.

International Adjustments

International adjustments under the international gold standard

In an earlier chapter, "Gold and the Price Level," some indication was given of the manner in which, under the gold standard, gold flows more or less automatically corrected international maladjustments. It is necessary, however, to examine this process in more detail. Let us, therefore, take the simplest possible case, with only two countries, both on the gold

standard, so that gold flows are reflected directly in the monetary system and price levels of each. In the interest of simplicity all capital movements are also excluded, leaving for consideration only the movements of goods and services and of gold.

Suppose that country A begins exporting more goods and services to B than it is buying from B. This means an increased demand by importers in B for A's currency in order to pay for this excess of A's exports. Gradually the price of A's currency will rise in terms of B's currency until the gold export point is reached. Thereafter gold will be shipped from B to A to cover the export excess from A, replacing any further shift in exchange rates. The unbalance in trade has set up a gold flow. How does this gold flow act to correct the unbalance and restore a rough sort of equilibrium in trade between the countries?

In country A, there is now more gold, which will either be used by the government as partial backing for new currency issues or go into the banking system to serve as reserves for credit expansion. In either case the total quantity of money and credit will be increased. Other things being equal, this will mean higher prices and higher costs for goods produced in country A. It will also give citizens in A greater incomes to spend on goods produced in B.

In country B, the reverse will be true. The outflow of gold necessitates monetary and credit contraction, since the government loses part of its currency backing, and/or banks lose part of the reserves on which credit had been expanded. Other things being equal, this diminished quantity of money and credit will mean lower prices and lower costs for goods produced in B. Also citizens in B will have lower incomes to spend on products produced in A.

As a result of these changes, country A will begin buying more than before from B (because of the lower costs and prices² of goods in B and because of increased purchasing power over B's products), while B will purchase less than before from A (because of the higher costs and prices of goods in A, and because of decreased purchasing power over A's products). Thus the flow of gold from B to A will be checked and an equilibrium gradually reached between the two countries in which their trade is roughly in balance, with gold used only to settle temporary cases of unbalance.

With three or a dozen or any other number of countries, the method in which gold flows lead to equilibrium in international trade has been essentially the same. When nations are on the gold standard, gold flows automatically lead toward international equilibrium through the monetary systems of the countries involved. Gold flows will not long continue in the same direction without automatically bringing about a check to

² Strictly, prices of internationally traded goods must be identical in all trading countries, transportation costs excluded. Therefore, B sellers will obtain a larger share of the identical price market than before because of lower costs, rather than because of a lower price than that received by A sellers. However, the looser statements above are simpler and sufficiently accurate for elementary purposes.

themselves by raising prices, costs, and incomes in the gold-receiving nation. But gold flows tend to bring equilibrium only so long as they are permitted to affect prices, costs and incomes in the countries involved.

Although this picture is much oversimplified, it represents loosely the process of adjustment in world trade during the pre-1914 gold standard period. When the United States received more gold from abroad or when newly mined gold was brought to the mint, the metal did serve as a basis for new money and it did provide new bank reserves on which banks usually expanded credit. When we lost gold, either through international trade or through having it withdrawn from circulation for hoarding in times of alarm, the amount of money and credit was diminished. Other countries ordinarily followed more or less the same procedure, though both here and abroad some steps were taken from time to time to affect the price level aside from the influence of gold flows. It was when countries became increasingly unwilling to have domestic monetary and fiscal policies primarily controlled by gold flows that the international gold standard gradually broke down and gave way to other arrangements.

International adjustments under fluctuating exchanges

More or less automatic correctives to any unbalance existed under the international gold standard, so that payments for goods and services plus capital movements into a country roughly equaled payments plus capital movements out of the country, with only temporary gold movements. How, if at all, does this international adjustment take place when the international gold standard is absent?

Assume an international maladjustment so that the payments being made to country A by country B greatly exceed those of A to B, again treating only commodity items in international exchange and limiting the discussion to two countries. This disequilibrium means that A is selling more to B than B sells to A. But equilibrium will again be restored. As people in B try to obtain foreign exchange to make payments in A, the price of A's currency (let us say, dollars) will go up in terms of B's (let us say, francs). As the exchange rate moves against the franc, it takes more francs than previously to buy a given amount of goods in A and, conversely, fewer dollars than previously to buy a given amount of goods in B. In addition, the incomes of people in A will have been increased by increased sales to B, while the incomes of people in B will have been decreased because of lower sales to A.

Since A citizens now are richer and find it easier to buy in B, B's exports will increase, while exports from A to B will decrease for the opposite reasons. Thus, as people in A buy more from B and people in B buy less from A, their exports and imports tend to return to equality; and as long as the greater payments of people in B continue to raise the price of dollars relative to francs, this tendency toward equilibrium will continue. The change in the dollar-franc exchange rate as a result of the greater payments originally made from B to A has led to an equilibrium of goods bought and sold by the two countries, just as did the flow of gold under

the international gold standard. *Under either arrangement* international adjustments occur more or less automatically, though sometimes rather slowly and painfully.

But there are important differences between the two methods of adjustment that require attention in considering whether it will after the war be desirable to attempt to return to the international gold standard, to peg exchange rates by international agreements without the gold standard, or to permit fluctuating exchanges. As has already been noted, the greater certainty and convenience provided international trade and finance by exchange stability is of major importance. On the other hand, it was a serious drawback that the gold standard tied the supply of money and credit of participating countries rigidly to gold flows. Especially are such gold-induced price and cost changes highly undesirable because they are likely to precipitate booms or depressions as prices move up or down.

By contrast, international adjustment under fluctuating exchanges does not necessitate such a complete internal readjustment. With exchange rate shifts, the adjustment is largely centered in the prices and output of internationally traded goods. In the preceding example, when B started buying less of A's goods as the price of dollars rose in terms of francs, this directly checked production in the export industries of A. Of course such a contraction in A's export industries would have some deflationary effect on other industries. But the seriousness of this effect would be much less than under the gold standard, where the outflow of gold would necessitate a contraction of money and credit, pressing down on the whole price level while sticky costs remained high. Under the gold standard, general deflationary pressure on the whole economy would result in A; while in the adjustment by fluctuating exchanges the contraction because of decreased exports would fall directly on the export industries (where it ultimately must fall in any case), thus avoiding to some extent general deflationary pressure. Just the converse would be true in the case of the opposite adjustment—with flexible exchanges there would be less likelihood of a general boom since expansion would be centered largely in the export industries. •

This advantage of the flexible exchange method in avoiding unnecessary deflationary pressure during international adjustments is of great importance, given the susceptibility of modern economic systems to booms and depressions. Especially is it important when large short-term capital "flights" are likely to occur as a result of unsettled world political and economic conditions. Under the gold standard, tremendous expansionary and deflationary pressures would have to be exercised to bring about adjustment to such drastic and unpredictable gold flows. No economy could stand the shock of such expansions and contractions without severe booms and depressions. On the other hand, the importance of stable exchange rates to international trade and finance is not to be under-emphasized. A widely favored compromise proposes domestic monetary stabilization (of the sort discussed in an earlier chapter) to mitigate business fluctuations, coupled with international cooperative exchange

controls to minimize exchange fluctuations—thus obtaining the major advantages of both relatively stable exchanges and domestic business stabilization. But without some assumption as to the future of world political and economic conditions, it is impossible to say what future policy should be. Further, what happens to the numerous barriers to international trade and investment discussed in the next chapters must play a large part in influencing postwar plans. At the conclusion of those chapters alternative courses of economic policy in the international economy can be more fully weighed.

CHAPTER 54

Interferences with International Trade: The Tariff

The Background of the Tariff

DURING the 1600's an economic doctrine called "mercantilism" became prevalent. One important tenet of this doctrine was that a primary aim of government policy should be to get more gold and silver, especially from abroad. Unless treasure could be obtained simply by seizing it in the new world, a "favorable balance of trade" was required to get gold. If a country exported more than it imported, it received gold for the balance. There developed the strong feeling that a "favorable balance of trade," which brings into a country gold and silver, should be a basic goal of foreign trade.

This preoccupation with gold and silver as major ends of international trade has continued up to the present in the minds of probably a majority of people. Added to this venerable argument for a "favorable balance of trade" have been numerous others: that a "favorable balance of trade" increases domestic employment, that it increases domestic wages and raises the plane of living, that it makes the nation self-sufficient, and so on. All these arguments claim advantages from exporting more than we import—all spring fundamentally from the old favorable balance of trade position. Although economists for at least a century and a half have been pointing out the fallacies in most of these "favorable balance of trade" arguments, only with the huge inflow of gold during the 1930's did the general public become uncomfortably aware that there is such a thing as getting more gold than we want, and getting it by giving to foreigners in exchange real goods, services, and investments in this country. The tremendous material needs of war in the last few years have further emphasized the importance of real goods and services as compared with gold and money flows. None the less, there is little real evidence that the hold of most of the other favorable balance of trade arguments on the public mind has been substantially weakened, in spite of numerous war moratoria on trade restrictions among allies.

It is important to examine the policies in which this "favorable balance of trade" doctrine has been reflected in more recent years, or, in other words, to examine the various means by which nations have interfered with free international trade in attempting to improve their own welfares. If the favorable balance of trade doctrine which postulates that we gain

by having exports exceed imports is fallacious, why has it survived so long and so greatly influenced national policies in almost every nation?

Development of the Tariff

By far the commonest method of interfering with international trade in attempting to promote national advantage over the last century has been the tariff, although since the First World War other methods have greatly increased in importance. A tariff is a tax on imports into a country,¹ either so much per article (called a "specific" tariff duty) or so much per dollar value of imports (called an "ad valorem" tariff duty). Two effects of such a tax become obvious immediately: (1) it tends to shut out imports, and (2) it yields revenue on those articles which come in in spite of the tariff. But the two results are incompatible—the more the tariff keeps out imports, the less revenue it raises; the more revenue it raises, the less it shuts out imports.

In this country up to the Civil War the tariff was the primary source of revenue of the federal government. But from 1816 on the tariff was increasingly used for "protection" as well as for revenue, as particular interests within the country sought protection against foreign competition. Following the Civil War the "protective" aspects of the tariff became steadily more important and the "fiscal" aspect less important, until in the present century the former consideration has dominated completely. Although at the beginning of the present war tariff duties still yielded several hundred million dollars annually (\$319,000,000 in 1939), decisions to increase or decrease tariff rates have long depended almost exclusively on the issue of protection to domestic interests, with the revenue consideration largely incidental.

It is difficult to trace simply the growth of tariffs because of the great complexity of tariff regulations. The trend has been heavily toward increased protection over the last century, and this trend culminated in the Smoot-Hawley Tariff of 1930, which imposed the highest barriers yet seen.² Merely a listing of the commodities covered by this act with rates of duty takes 212 pages of fine type on 8½" by 11" paper.³ Almost half of the commodities listed have effective duties exceeding 100 per cent of their value before payment of the duty. And of all the commodities listed, less than one out of ten is admitted free of duty—these free items are largely those whose production is almost impossible in the United

¹ In some cases special taxes on exports have also been termed "tariffs," as for example the long-standing British tariff system on exports from the twelfth to the eighteenth centuries.

² During the 1930's the Reciprocal Trade Program, directed by Secretary of State Hull with President Roosevelt's strong support, made some headway in reducing the tariff barriers on many commodities through reciprocal agreements with other countries to lower tariffs on particular products.

³ In the *Statistical Classification of Imports into the United States*, Bureau of Foreign and Domestic Commerce (Washington, 1937). Duties are as of Jan. 1, 1937.

States. As compared with this, the tariff acts of a century ago were simple and narrow in scope, even though the rates charged were in some cases almost as high. Growth in United States tariff barriers has come more from a widening of the list covered than from a raising of the average level of rates. Yet from about 17 per cent between 1812 and 1816, the average level of tariff duties has risen to around 50 per cent under the Smoot-Hawley tariff, a figure reduced to around 40 per cent by reductions under the Hull Reciprocal Trade Program. On several occasions tariff reductions have been promised by political parties, but only three times between the Civil War and the Second World War were appreciable reductions made, and the last of these was by administrative bargaining under the Reciprocal Trade Program rather than through Congress. Throughout this period the United States has been among the leaders in erecting tariff barriers; we cannot explain our growing tariffs as retaliation against foreign tariffs except in minor degree. Especially since the First World War the United States must accept a large share of the blame for setting under way international trade restrictions.

One common fallacy in regard to the importance of tariff barriers must be guarded against. Recent (prewar) statistics show that almost 65 per cent of our imports came in duty-free, and tariff advocates use this fact to argue that therefore the tariff did not seriously curtail trade. But this is a fallacious argument. United States tariff duties have been so high on large numbers of commodities that imports are completely excluded. The reason so large a *percentage* of imports have come in duty-free is that many products are largely excluded, making the few duty-free articles a high proportion of the remaining total imports.

With the spread of hostilities following 1939, war pressures became more and more dominant over international trade and finance. Even before the United States formally entered the war, Congress had passed the "Lend-Lease" provision, directing that war materials be sent to certain anti-Axis nations without monetary payment or promise of such payment; instead, after the war the receiving countries were to return the goods in kind, or some other goods determined by agreement with each country concerned. With entrance of the United States into the war, this provision was broadened still further, so that even the provision for eventual repayment in goods could be waived (a provision vastly different from that prevailing during the First World War when the United States acquired a heavy creditor status through sales of war necessities to our allies). With these developments the traditional mechanism of international trade has been discarded in many respects and many trade barriers between allied nations have been modified or suspended for the duration of the conflict. How permanent these changes are will be demonstrated only after the war. That a sharp division of opinion will occur appears certain, and consideration of the merits and disadvantages of peacetime trade restriction continues of major importance, even in the midst of the international chaos and realignments of war.

Effects of the Tariff

For well over a century the protective tariff has been the subject of violent political debate. The general lines of argument today are much the same as they were a century ago. Therefore, it is perhaps more useful to set forth the various arguments popularly advanced for tariff protection and to examine them one by one than to undertake a general formal analysis of the effects of tariff barriers. Consideration of these various popular arguments brings out the economic effects of trade restrictions, and permits the direct application of economic analysis to "real world" issues.

The "infant industry" argument

The first argument used in the United States for adoption of a protective tariff was that "infant industries" in this country should be protected from foreign competition until they had become firmly established and able to stand on their own feet. For example, in 1816 the newly founded iron industry found it very difficult to get established in the face of strong competition from well-established British iron makers. It was argued that if a protective tariff were established to shut out foreign competition for a few years, this would give the American iron industry a chance to develop to the point where it not only could meet, but could outstrip, foreign competitors. Iron would be available here at lower prices than from abroad. Further, the establishment of this new industry in the United States would give employment to American workers and improve business conditions here.

There is some validity to the "infant industry" argument. It is very difficult for a new industry to establish itself in the face of vigorous competition from older competitors, even though the new industry once established will be more efficient than the old. In the instance of iron, the protective tariff probably speeded the growth of an industry that would in any case have prospered. Perhaps the most significant fact is, however, that the tariff today remains on iron and steel products, and it would be difficult to call the iron and steel industry today an "infant." Historically, tariffs imposed to promote "infant industries" have almost never been removed after the industry has grown to "maturity." By the time the industry has grown so far as to make further use of the "infant industry" argument impossible, other arguments are found for continued protection.

Owing to the nature of tariff making in the United States, protection, once achieved, can almost invariably be retained. Congress passes tariff acts covering thousands of products, and most of the actual decisions are made by small committees in the House and Senate. Obviously no Congressman can know about more than a comparatively small number of these thousands. But each Congressman is pressed by his constituents to vote for protection for the goods that they produce. Thus, Congress-

men from Colorado are urged to vote for a tariff on sugar to protect sugar-beet raisers; Congressmen from New England are urged to vote for a tariff on shoes to protect New England shoe factories; and so on. When these items come up for consideration, each Congressman is interested in getting his own constituents protected, even though in principle he may be in favor of lowering tariffs. Each Congressman, in attempting to get votes for his particular rate increases, is almost certain to have to agree to vote for his fellows' duties in exchange for their votes in favor of the ones he wants. Thus, the Colorado Congressman agrees to vote for a tariff on shoes if the New England Congressman will vote for one on sugar. The tariff is the classic example of special group pressures at work in dictating government policy—each group interested in protecting its own commodity, but willing to help others get higher protection in return for votes for its own commodity. Such logrolling is by no means confined to tariff-making—it is only especially flagrant there—but it has gone far to ensure that protection given infant industries will not be removed.

The infant industry argument is valid on two conditions: (1) that the infant industry actually gives promise of developing into a self-supporting mature industry, based on a comparative cost advantage; and (2) that the protective tariff be removed as soon as the industry grows out of the infant stage. Although the first condition may be fulfilled in numerous cases, the second never seems to be. Therefore, economists are loath to support tariffs for infant industries because removal of these tariffs seems to be almost impossible. Moreover, any infant industry that has an important advantage in efficiency will gradually develop even without a protective tariff, though more slowly than with protection.

The "national self-sufficiency" argument

When fear of war exists, there is a strong argument for making a nation economically as self-sufficient as possible, especially if potential enemies are likely to control the seas. The great dependence of the United States on the Far East for tin and rubber supplies at the beginning of the present war clearly indicates the possible importance of this consideration. In such a situation, the desire for economic self-sufficiency may overbalance all other considerations and a country may well decide to erect tariff barriers to stimulate domestic production of vital products, regardless of what the costs of such economic self-sufficiency may be.

Rubber is a good example. It has been demonstrated that rubber can be produced synthetically in the United States, but at a much higher cost than that of the natural rubber. With free trade we should produce no rubber; instead we should be specializing in those commodities where our comparative advantage is greatest. If before the war we had ceased buying rubber abroad, we should have had to divert productive agents from producing, let us say, automobiles (where their productivity was high) to producing synthetic rubber (where their productivity was lower). Self-sufficiency necessitates a reallocation of resources away from their most "efficient" into less "efficient" productive uses and so lowers the

national plane of living in peacetime, and this cost in terms of decreased per capita real income may be a heavy one. Yet if it spells the difference between victory and defeat in an on-coming war the price may be a small one indeed.

The economist cannot say whether or not a nation should attempt to become partially or wholly self-sufficient. This decision involves weighing the political, social, and economic advantages of self-sufficiency (economic nationalism) against the political, social, and economic disadvantages of such a policy. The economist's function is to point out what self-sufficiency means in terms of poorer allocation of resources and lowered planes of living, in order that this factor may be weighed with others in deciding whether or not more self-sufficiency should be sought.

The "favorable balance of trade" argument

Most naïve of all the arguments advanced in support of the tariff is the desire for a favorable balance of trade for its own sake. There are two major fallacies involved. The first is that there is something inherently favorable about a "favorable" balance of trade. A continued "favorable" balance means that a nation continually gives foreigners more goods and services than they give it, receiving gold in exchange. The argument that a nation ought continually to have such a trade balance loses sight of the basic purpose of trade—namely, to better one's own position by exchanging goods owned for other goods desired more. The gold received through a favorable balance of trade is useful as a means of procuring other goods and services when respent in foreign countries, or it may be useful as a backing for the domestic monetary and credit structure. But when the gold is held beyond this latter need instead of being respent in foreign trade, the hoarding country finds itself very much in the position of King Midas, who became very hungry when everything he touched turned to gold. By giving away real goods and services in exchange for gold, which is then simply held idle (at considerable storage expense), a nation is clearly the loser from a continuously favorable balance of trade. Yet this is exactly the result obtained when the favorable balance of trade argument is carried to its logical conclusion.

Even this is not the extent of the fallacious reasoning in the "favorable balance of trade" argument. The argument runs that the United States should continually sell more to foreigners than they sell to us. But this is impossible as a long-run phenomenon. The only way that foreigners can buy from us is by somehow having funds to pay for their purchases. The way that they get dollars to buy American goods is by selling their goods to us. For some time they can continue to buy more from us than they sell us, paying for the balance in gold or by giving us investments in their industries, but eventually it is only by buying from foreigners that we can expect to sell to them. A continuous "favorable balance of trade" is neither favorable nor possible, except where the paying countries actually mine the gold that they use to pay for their unfavorable balances.

By shutting out foreign products in the attempt to secure a favorable

balance of trade, therefore, the tariff involves a direct loss to the imposing country's export industries and an indirect loss to all persons receiving incomes through them. The less that we are willing to buy from foreign nations, the less they can buy from our export industries, a group whose importance is indicated by the tables presented in Chapter 52. The loss involved to and through these industries from imposition of tariff duties is only one of the costs of tariffs; correlative losses to consumers will become apparent as other tariff arguments are considered.

Yet, in spite of this strong condemnation of the "favorable balance of trade" argument for protective tariffs, there is a certain element of truth in the argument as a short-run proposition. For a short period of time, a country *may* increase employment, production, and income by raising tariffs to create temporarily a favorable balance of trade. This possibility will be analyzed presently under the "increased employment" argument. But this gain from a temporary favorable balance of trade is strictly a temporary, short-run advantage, which *may* be of aid especially in times of depression—it must not be mistaken to mean that a long-run favorable balance of trade is either ordinarily possible or desirable.

The "protect home industry" argument

Aside from the self-sufficiency argument, the basis for most popular arguments advocating tariff duties is that we ought to protect home industry against low-cost foreign competition. There are many versions of this "protect home industry" argument, but all have essentially the same reasoning. Prefatory to considering these various versions, let us examine the basic claim that home industry should be protected, for the fallacy involved shows up most clearly in this direct contention.

A domestic industry asking for a protective tariff on its product argues that unless it receives such protection its market will be lost to foreign competitors, thus forcing the domestic industry out of business, throwing workers out of jobs, and generally decreasing national welfare. Actually in many cases industries seeking protective tariffs would be able to retain their markets without tariffs, and want tariffs merely to enable them to raise domestic prices monopolistically without fear of foreign competition. Indeed, the tariff has been called "the mother of monopolies" since it permits domestic monopolies to exist without fear of foreign competition. There can be little justification for tariff protection for these industries. But suppose that the industry actually must receive protection or it will lose much of its market to foreign competitors. What will be the effects of giving protection to this industry, say sugar-beet production?

The first and most obvious effect is that consumers of sugar must pay more for sugar than if free import had been permitted. If sugar comes in over the tariff, consumers must pay the regular price of the foreign sugar plus the tariff duty. If the tariff succeeds in shutting out foreign sugar, consumers must pay a higher price for domestic sugar since domestic producers cannot, without a tariff, produce and sell at a price low enough to meet foreign competition.

The second effect is that domestic sugar producers are subsidized by the tariff—they are permitted to charge higher prices than would have been possible had not foreign competition been shut out. Either this protection will permit them to make profits where otherwise profits would have been eliminated by foreign competition, or it will lead to overinvestment in the industry with costs rising to the artificially maintained price, or it will permit producers to stay in business without loss where otherwise foreign competition would have eliminated them. Clearly domestic producers are the great gainers from a tariff on their product.

In effect, therefore, a tariff on sugar is a subsidy to domestic sugar producers, financed by consumers through increased sugar prices. Moreover, the more inefficient domestic producers are, the higher is the tariff that they need to protect them against foreign competition and the larger the subsidy that they receive from such protection. Indeed, usually the industry's argument for a tariff is on precisely this ground—that the tariff should be high enough to equalize domestic and foreign costs of production. A suggestion in Congress that special taxes be levied on consumers to finance subsidies to producers, each subsidy to be based on producers' inefficiency, would horrify everyone—yet this is the result of a protective tariff designed to protect home industry.

Data are available to show this effect. Continuing the example of sugar—during the 1920's the tariff on Cuban sugar (our main sugar import source) was 1.76 cents per pound on raw sugar, 1.89 cents per pound on refined sugar. In 1928, about 50 per cent of our sugar came from Cuba, 32 per cent from our island possessions, and only 18 per cent from domestic cane and beet sugar producers. As a result of the tariff the domestic price of sugar was higher than the international price by almost exactly the amount of the tariff. Because of this higher price permitted by the tariff, domestic producers received approximately a \$43,000,000 benefit. But since the price of imported sugar was also higher by the amount of the tariff, the higher prices constituted a special burden of approximately \$289,000,000 on consumers. Since the average family consumption of sugar that year was about 415 pounds, the higher price of sugar because of the tariff cost the average family about \$9.

Consumers were thus the great losers from the tariff, as well as American export industries whose sales to Cuba were restricted by our refusal to buy Cuban sugar. Domestic sugar producers gained, but much less than the burden on consumers. Since sugar from our island possessions comes in free of tariff, producers there also gained. The government collected about \$135,000,000 duty on Cuban sugar, but this was less than half the burden on sugar consumers, and an excise on sugar is hardly an equitable basis for taxation since most people consider sugar a necessity. Thus, the cost of the tariff to consumers was far greater than the subsidy to domestic producers, and even greater than the subsidy plus the revenue to the government, because island possession producers also received the subsidy of higher prices. Indeed, it has been estimated that if we were to remove the tariff and get our sugar correspondingly cheaper, sugar

consumers could afford to give outright a living wage to each regular employee in sugar-beet raising for the rest of his life and still be considerably better off than with the tariff. Actually, however, as we shall see, removal of the tariff would probably not make these workers public charges for life. Although it certainly would throw most of them out of jobs in raising sugar beets, the likelihood is that they would be re-employed elsewhere at higher wages. But this extreme example of subsidy for life indicates how costly and inefficient a tariff may be.

In 1930 the tariff on sugar, and almost all other articles, was raised even higher. Thus 1928 is not the most extreme case which might have been taken. And although sugar is among the most indefensible tariffs that we have, it is a good example of the results of a high protective tariff. Another example in the same year was the tariff on wool and close wool substitutes, which cost American consumers somewhere around \$130,000,000. The benefit to domestic wool producers, who supply a relatively small part of our consumption, was about \$43,000,000. These figures on sugar and wool are only rough estimates. Yet they are accurate enough to demonstrate clearly the effects of such tariffs.

But the tariff advocate is not likely to be stopped by these considerations. Evading the fact that the tariff is essentially a subsidy to inefficiency paid by consumers, he is likely to argue that if we were to remove tariffs we should lose our favorable balance of trade, domestic industries would fail, workers would be unemployed, wages would be depressed, and the plane of living would fall. These are to some extent separate arguments presenting specific claims; each deserves careful consideration.

The "equalize costs of production" argument

Many persons have argued specifically that we ought to levy a protective tariff just high enough to "equalize costs of production here and abroad," no more and no less. Advocates of this proposal argue that this would not subsidize inefficient American industries, but would simply permit them to compete with low-cost, low-wage foreign production "on equal terms." Let us postpone temporarily the argument that we need a protective tariff to protect our high wages against low foreign wage competition, and see what is the merit of this proposal that tariffs shall be set just high enough to equalize domestic and foreign costs of production, which has, as a matter of fact, been written into our law. Certain portions of the Smoot-Hawley Tariff Act of 1930 are stated in exactly these terms.

The first question to be asked is, whose domestic costs are to be equalized with whose foreign costs? If the costs of the most inefficient domestic producers are taken and the tariff set high enough to provide full protection, the rate obviously gives a special subsidy to all more efficient producers in the industry. If the costs of the most efficient producers are taken, all others complain that the rates set are not high enough to protect them against the low foreign costs. If some sort of average is taken, there

results a combination of obvious subsidies with "insufficient protection." Similarly, there is usually a great variety of foreign costs, both as between different countries and as between different producers in the same country. Which of these shall be considered the foreign cost to be equalized? If the highest foreign cost is taken, then lower cost producers can send in some products over the tariff. If the lowest foreign cost is taken, this sets up a complete barrier against all foreign producers, a result that would probably be satisfactory to domestic producers but hardly conducive to the welfare of domestic consumers.

But let us disregard for the moment these problems of what costs to equalize and ask what would be the effect of a tariff that equalized domestic and foreign costs of production on all products? The answer may be somewhat shocking: *If all costs were so equalized there would be no foreign trade whatsoever.* Foreign trade is profitable when there is a sufficient differential in costs to make possible the payment of transportation charges and sale at a profit in a foreign market. If tariffs were high enough exactly to equalize costs of production, it would not be profitable to ship any goods, since they could not be profitably sold at a price low enough to meet the price of domestic producers who avoid international shipping charges.⁴ American exports could not compete in foreign markets; foreign sellers could not compete here.

Some implications of adjustment to such a "self-sufficient" economy have been indicated in a previous section. The major long-run result would be a mass shift of resources from export industries to the production of goods previously imported. It is estimated that in agriculture, elimination of our prewar foreign markets would mean that we should have to retire about 22,000,000 acres of cotton land, about 9,000,000 acres of wheat land, about 665,000 acres of tobacco land, and about 9,000,000 acres of corn land now used for feeding hogs for export, to cite only some of the more important products. In total, loss of farm export markets would necessitate the retirement of possibly 40,000,000 acres of average farm land. This land today supports a population of some 3,000,000 people. Even if the employment possibilities in industries supplying previously imported products should increase correspondingly, the problems of transition would be tremendous for human beings, and what use could be made of the abandoned land is very doubtful.

In the industrial field, the number directly dependent on exports for a living is much larger than in agriculture, though such workers are likely to be more mobile, and hence their transfer to new occupations might not present quite such formidable difficulties. Rough estimates for the immediate prewar period place the number in industry directly dependent on foreign sales at around 7,500,000 (including employee dependents),

⁴ This assumes that domestic transportation costs are lower than those between nations. This assumption is not always true, but this is a refinement that is not essential at the elementary level since no new principles are introduced by complicating the assumption

making a total of over 10,000,000 directly dependent on foreign trade at the beginning of the war. Accompanying the human problems of such a mass migration away from export industries would be the problems connected with retiring some three billion dollars worth of agricultural land and rendering worthless perhaps nine billion dollars worth of industrial plant and equipment whose capacity would be superfluous for the domestic market alone. Thus, quite aside from the economic costs of national self-sufficiency, noted in previous and succeeding sections, the problems of transition imposed by a prohibitive tariff barrier are in themselves terrific. The logical consequences of high-sounding proposals to "equalize foreign and domestic costs to give the domestic producer a fair chance" would be quite different from what the title of the proposal connotes.

The "protect our high wages and high plane of living" argument

Especially in the United States a major argument for the tariff has been that it protects high domestic wages and the high American plane of living. In other countries wages and planes of living are typically much lower than here; therefore it is argued that unless we have a protective tariff, international competition will push down high American wages to the level of foreign wages. With the decline in wages will inevitably go a fall in planes of living.

To analyze this argument, we must ask first why wages are now generally much higher in the United States than in most other countries. Since in the long run the wage paid can be only as high as the marginal contribution of the worker to the sales revenue of his employer, the worker's marginal productivity sets a rough ceiling for his wage. Where competition for workers exists among employers and where trade-unions are present to prevent exploitation, wages tend toward this level, although under conditions less favorable to the worker they may remain substantially below it. Fundamentally, high American wages rest upon the high productivity of American workers—because of their skill and training, because of the abundant natural resources with which they have to work, because of the efficiency of American mass production methods, and so on. Anything that raises the productivity of American workers makes it possible for them to receive higher wages. And each worker can receive his highest possible wage when he is working in that industry where his marginal productivity is highest. As emphasized in an earlier chapter, this allocation of resources to those industries where their marginal productivities are greatest will be encouraged under free trade, wherein each country specializes in production of those goods in which it has the greatest comparative advantage.

Compare this situation with that after a tariff has been passed to "protect" high American wages. The tariff will mean that relatively inefficient American industries will grow up where otherwise they could not have existed. These relatively inefficient industries will of course employ workers and other resources. But at the same time, American

exports will fall off because foreign countries cannot long buy from us unless we buy from them. Therefore fewer workers and other resources can be employed in the export industries. The net result will be a shifting of workers and other resources out of highly efficient export industries (where they would be with free trade) into less efficient protected industries. As workers move from high marginal productivity positions in export industries to lower marginal productivity positions in protected industries, the wages they can receive in the new jobs are necessarily lower than they could be before the shift. *The long-run result of a protective tariff is to lower real wages, not to raise them.*

The same reasoning is true for the return to other productive agents. The highest payment any productive agent can receive is set by its marginal productivity. Imposition of a tariff that protects inefficient industries and cuts down sales to foreign buyers causes a shift of productive agents to these inefficient industries where agents' marginal productivities are lower than under free trade. For all sorts of productive agents, payments tend to be lower, not higher, as a result of tariffs intended to "protect" high American wages.

It follows, therefore, that a protective tariff lowers the national scale of living, not raises it. A nation achieves its highest plane of living with productive resources allocated where their marginal productivities are highest. The tariff, by causing a less favorable allocation of resources, lowers the scale of living that a country can have.

To illustrate this point, let us take perhaps the most extreme example cited by proponents of a tariff to protect American wages and the scale of living. Since Japanese labor is paid perhaps 25 cents per day, it was long argued that if we had permitted free import of Japanese goods this would surely force American workers to take much lower wages. In considering this argument, one needs first to see that even though Japanese daily wages are much lower, it is not necessarily true that labor costs of Japanese producers (say of light bulbs) are correspondingly lower. The Japanese worker gets much lower wages, but he also produces far less in one day. Therefore the labor cost per light bulb might be as much for the Japanese producer as for the American in spite of the great daily wage differential. If, for example, the daily output per worker in Japan was 10 bulbs, the labor cost per bulb would be 2.5 cents. If the daily output per American worker was 200 bulbs and the daily wage \$4.00, the labor cost per bulb would be only 2 cents. The wage cost that counts is the wage cost per unit produced.

But this by no means answers the tariff argument, for it probably is true that labor costs on many articles are lower in Japan in spite of the higher output per worker here. Suppose, to continue the hypothetical example, that Japan could far undersell us on light bulbs, and we had removed the tariff on light bulbs. This of course would have meant that workers in the light-bulb industry would have been thrown out of jobs. But since Japan would then be selling ~~more~~ to us, she would have had the power to buy more from us. She would buy from us those goods that

we were able to produce most efficiently and cheaply (let us say, automobiles). Workers unemployed by elimination of the light-bulb industry would gradually tend to be drawn into the automobile industry. But since the United States is, by assumption, comparatively more efficient in producing automobiles than light bulbs, workers can receive higher wages making automobiles than making light bulbs. Although light-bulb workers would surely be temporarily unemployed and, if personally unadaptable, might never be re-employed, the general public would clearly gain by obtaining cheaper light bulbs. Moreover, gradually more workers would be drawn into the high productivity automobile industry where wages are highest. That removal of an existing tariff may injure protected producer groups is undeniable; yet a net gain in plane of living accrues to the whole economy.

It is true, however, that if American laborers have received the benefit of minimum-wage or maximum-hour provisions, of social insurance, or of other social reform steps increasing the effective labor-cost per unit of output, elimination of American tariffs would make it harder for American labor to maintain these advantages. Such labor standards, or other artificial steps to raise wages, would put American producers at a relative disadvantage in comparison with lower standard foreign labor. Free trade would therefore tend to force elimination of the higher standards or to induce lower wages to offset these additional costs, especially as other productive agents were substituted for the relatively expensive American labor.

The "increased employment" argument

In the long run the effect of a protective tariff on employment is neutral—it neither increases nor decreases employment, but merely shifts resources from one group of "more efficient" industries to another group of "less efficient" industries. However, the short-run adjustments in moving toward the new long-run equilibrium may be very important. In the previous example of removing a tariff on light bulbs, American workers thrown out of work there could eventually be reabsorbed elsewhere, possibly at higher wages, and surely the general level of American real wages would rise. But this process of shifting from one job to another may be very slow and painful. Removing a tariff necessarily results in temporary unemployment, unemployment that may persist stubbornly under certain circumstances.

If workers were easily mobile between different jobs and between different locations, such unemployment would be only a minor problem. But actually immobility is very great, especially where considerable retraining is required for the new job. The difficulties of shifting to new jobs may be further increased by restrictions on entrance to the new occupation, such as strict union apprenticeship rules and high entrance fees and dues, or by monopoly restrictions exercised by employers. It is of little consolation to a light-bulb maker to know that eventually his wage may be higher if in the meantime he is unemployed, his children

underfed and ill-clothed, his home and insurance policy forfeited. Though he benefits from the lower cost of light bulbs as a consumer, his producer interest is dominant, and he as *an individual* may never be able to shift to making automobiles. If tariffs are to be removed or seriously reduced, the reduction must be handled carefully to minimize undesirable short-run effects. During good times with full employment, removal of tariffs could probably cause tolerably little painful readjustment because of the comparative ease of finding new jobs in such periods. But in times of depression and unemployment, removal of tariffs would probably be unwise except slowly and in special cases. In whatever period, adequate government relief to the tariff-unemployed and government aid in retraining such persons for new jobs could go far to mitigate any short-run disruptions that might result.

As removal of existing tariffs may cause temporary unemployment, so imposition of new tariffs during depression may temporarily help solve a nation's domestic unemployment problem. In the long run a new tariff merely results in a *shift* of resources into less efficient industries, but if the tariff is imposed during a period of unemployment, it may temporarily stimulate the newly protected domestic industry and put idle resources to work. During a severe depression, even such a temporary increase in employment may look very attractive, in spite of its long-run disadvantages.

Actually, however, this short-run increase in employment from imposing new tariffs is likely never to materialize. Especially during a depression period, imposition of new American tariffs would almost certainly be the signal for retaliatory tariff increases by other nations, reducing American exports as much as the increase in domestic production from increased protection. It was exactly by this process of retaliatory restrictions that trade barriers were forced up so far during the 1930's. Each country sought to gain a temporary advantage by raising tariffs, and the net result was much higher tariff walls between nations without any of them gaining the desired short-run advantage. For illusory short-run gains, nations imposed long-run burdens on themselves and on each other; and once a protective tariff is imposed the process of removing it is at best slow and doubtful.

The "national economic stability" argument

During the nineteenth century, economists devoted the major portion of their attention to finding ways to raise the national plane of living. Relatively little attention was given to the problem of economic stabilization. "Progress," not security, was the goal.

More recently, emphasis has shifted strongly toward security as the major goal of economic policy, at the cost of economic "progress" if need be. Insofar as it refers to international trade policy, this shift has led to increased support for economic nationalism as a device for "insulating" the domestic economy against disturbances originating without. In some versions the argument is carried to the extreme of advocating complete cessation of foreign trade in the hope of stabilizing our domestic economy

at a level of reasonably full employment, even though the general plane of living thus attained would admittedly be far lower than that to which we have been accustomed in reasonably good times. In less extreme versions, advocates of the "security" approach suggest the need for an independent stabilizing monetary policy, possibly supplemented to some extent by new trade barriers as employment stimulants in depression periods.

It is, of course, true that participation in international trade renders a nation more susceptible to foreign economic disturbances than would be true under isolation. In particular, business fluctuations spread from nation to nation in part through the channels of international trade. As depressed countries decrease their purchases abroad, deflationary effects are felt in other nations; and conversely in boom periods. Yet withdrawal from all international trade to avoid such international repercussions, especially if domestic cycles may occur even without international trade, seems to many writers to be virtually killing the patient to cure him. Even such less drastic expedients as depression-period tariff increases and export subsidies stand on shaky economic ground in view of their dependence on the failure of foreign nations to retaliate if even short-run benefits are to be derived.

Actually most international cyclical disturbances appear to be transmitted at least as much through the monetary mechanism of exchange as through trade itself. The manner in which domestic monetary conditions often aggravated business fluctuations under the gold standard has been discussed. Most disastrous were periods like that following 1929, when fear and uncertainty caused violent shifts of short-term capital from one country to another. Under the rigid exchanges and monetary arrangements of the gold standard, such shifts exerted violent upward or downward pressure on the economies affected—sufficiently so that few nations were willing or able to cling to the gold standard. With continued international political instability and increasing emphasis on the powers of domestic monetary and fiscal policy to achieve business stabilization at high levels of employment, growing sentiment has appeared for complete independence of domestic monetary action. Exchange stability is viewed as a secondary consideration since international trade is adjudged as subsidiary to domestic stabilization. By following a domestic fiscal policy aimed exclusively or primarily at domestic stabilization, it is hoped that domestic security and stability may be obtained—if international trade and exchange conditions suffer, this is a price well worth paying, according to this point of view.

Such programs of domestic monetary stabilization are closely related to "self-sufficiency" tariff policies, and in fact are at the core of most rational arguments for "insulating" the domestic economy from foreign disturbances. Merely shutting off international trade in itself solves few problems, and generally, to the extent that independent domestic stabilization is sought, insulating action through monetary and fiscal policy seems a more effective and less costly approach. Other economists, how-

ever, question whether any independent domestic stabilization can succeed in the modern world, and prefer to place primary reliance on concerted international action to maintain the volume of foreign trade and the level of employment within each country.

There is one clear case where high tariffs may help substantially in stabilizing domestic business activity. This case occurs when a nation is highly specialized in producing one or a very few major products; Cuba's sugar economy and Brazil's coffee economy are outstanding examples. If foreign markets weaken for such one-product economy, the effect on the entire economy may be devastating. For such countries greater diversification may be an advisable step in order to lessen this dependence on foreign conditions, even though the industries stimulated are relatively inefficient. For countries already widely diversified, such as the United States, this argument has little significance.

Conclusion

From the consideration of these various arguments a summary of the most important economic effects of protective tariffs may be drawn. *In the long run*, a protective tariff necessarily lowers real wages and the plane of living, diverting resources from self-sustaining export industries to protected domestic-consumption industries and forcing consumers to pay higher prices. None the less, such tariffs may be desirable if national self-sufficiency is considered of primary importance. *In the short run*, advantages may be gained from imposing new tariffs to aid infant industries or to stimulate employment, production, and income during depression periods. However, infant industry tariffs are seldom later removed, and new tariffs can temporarily increase employment and income only if almost inevitable retaliatory steps are not taken by other nations. Lastly, history demonstrates dramatically how use of the tariff as an employment stimulant works strongly toward long-run restriction of trade and toward cumulative international friction. A variety of special cases can be postulated where peculiar combinations of circumstances would make it possible to obtain beneficial national results through imposition of tariff restrictions. But for the most part these are artificial, academic examples, and the basic economic case for free trade remains fundamentally unimpaired.

Tariff Making in the United States

If the protective tariff is open to such serious criticism and benefits only small groups at the cost of the rest of society, why has the United States so long had high protective tariffs? For well over a century this country has had among the highest, and often the highest, tariff barriers in the world. The answer to this question is twofold.

First, as previously stated, there is probably no other economic subject on which public opinion has been so greatly influenced by self-seeking propaganda and on which the public has been so uninformed. Probably most people still believe in many, if not all, of the arguments for the tariff

just discussed. Groups seeking protection have effectively presented their propaganda in Congress and to the general public. There is some evidence that the recent Reciprocal Trade Program has enlightened public opinion somewhat on the advantages of freer international trade, but there is still little reason to hope that much of a dent has been made in the general protectionist sentiment in this country.

Second, owing to the law-making processes that produce tariff bills and to the inertia of consumers, it is possible for small groups seeking protection to trade votes and achieve much higher duties than would be possible each on its own merits. The logrolling and vote-trading that goes on in making tariffs has already been discussed.⁵ The tariff is the classic example of group pressures at work to achieve legislation benefiting small groups at the cost of the general public. With the tariff as with advertising, marketing laws, and the like, the consumers are likely to be unrepresented and forgotten, the great unorganized mass. There is no effective lobby in Washington representing general consumer interests, but the various groups seeking protection have strong, effective lobbies. The result is that legislation is enacted to favor the groups bringing most pressure to bear.

Experience indicates that the most hopeful channel for effective downward revision of the tariff is taking the job of rate-setting as far as possible out of Congressional politics and putting it in more impartial administrative hands. The experience of the last century and a quarter should be ample evidence that there is little hope through Congressional channels. On the other hand, the experience of the middle and late 1930's under the Reciprocal Trade Agreements Program showed that if the power to bargain for reciprocal tariff reductions were placed in competent administrative hands, results could be obtained.

Under the Trade Agreements Act of 1934, the President was given wide tariff-making powers, including the power to lower United States tariffs on specific commodities as much as 50 per cent if the other country most vitally interested in the product would make reciprocal concessions by lowering its tariffs on our exports to it. Under this arrangement 22 reciprocal trade agreements have been entered into, a large portion of them with other American countries, though the outbreak and spread of war following 1939 cut short this effort. A wide range of commodities is covered, and reductions in many cases have been substantial. Moreover, since the "most favored nation clause" was made a part of the agreements, concessions made in the reciprocal agreements were generalized to all countries (except those specifically excluded because of discriminatory action against United States' exports), even though the other countries had not made corresponding concessions to the United States. If, for example, we lowered our tariff on lumber in an agreement with Canada, this concession would also apply to lumber imports from all other countries, although in some cases (especially agricultural

⁵ In connection with the "infant industry" argument above.

products) the new lower rates were to apply only to certain quotas with any imports in excess of this specified amount to pay the old higher rates. This quota provision was used as a safeguard to American producers that the new lower rates would permit only limited foreign competition on the more advantageous terms. In accordance with the philosophy of the "most favored nation clause," it was generally left open to any nation to get a share of these quotas.

The "most favored nation clause," by extending treatment for all nations equal to that given the most favored nation, has long been one of the best devices in international relations to widen the scope of tariff reductions and to prevent discriminations and their resulting friction. It has, however, not been used by all nations and therefore has had only a limited amount of success. Actually the concessions extended under this clause to countries other than the one making the reciprocal agreement have usually been much less extensive than might be supposed. In the most limited case, the Canadian agreement, about 95 per cent of the total imports of the commodities on which we made reductions ordinarily come from Canada. Therefore, the "most favored nation clause" had only minor practical importance, although in most other cases its economic consequences have been greater.

The significance of the Reciprocal Trade Agreement Program, coupled with the "most favored nation clause," lay partly in the tariff reductions obtained and their extension through "most favored nation" treatment. However, of probably greater significance were (1) the reversal of the worldwide cumulative trend toward ever-higher trade barriers, and (2) the part played in cultivating more congenial political relations with the other nations involved, especially those in South and Central America. The Trade Agreements Program has meant a substantial amount to the United States in dollars and cents; probably it has meant vastly more to us in terms of international goodwill and future economic relations. If it had had no result other than to check the cumulative tariff increases at their 1934 levels in the 22 countries involved, it would have been a major achievement in American tariff-making. Today removal of trade barriers among the Allies as war measures, especially between the United States and England, holds forth a promise of sweeping reductions in peacetime barriers. Formulation of the peace will do much to determine the course of trade restrictions in the decades ahead.

CHAPTER 55

Interferences With International Trade: Quotas, Export Subsidies, and Exchange Control

PROTECTIVE tariffs have long been the primary means of controlling international trade so as to "protect" domestic industries and wages and to "give employment" to domestic workers. Since the First World War, however, a variety of new devices have been developed for these purposes, and during the depression of the 1930's these came into great prominence, in many cases overshadowing formal tariffs as restrictive measures. In this chapter these newer methods of control over international economic intercourse are discussed, concluding with a brief consideration of possible future economic policies in the international sphere.

Quotas and Allied Devices

A tariff protects the domestic market for a commodity by limiting the amount of foreign articles that may come in. For this same purpose "import quotas" are often set on specific products, effectively limiting foreign competition by prescribing the amount of the foreign commodity that may enter per year. The import quota gives results similar to those of the tariff in shutting out foreign commodities and permitting domestic producers to charge higher prices because of the absence of foreign competition. And, like the tariff, the import quota is a direct consumer-financed subsidy to domestic producers, because the limitation of foreign imports permits domestic producers to charge higher prices and diverts resources to less efficient industries. Usually tariff duties are also levied on the imports making up the quota, yielding government revenue and heightening the import barrier. The most important difference between a tariff and a quota is that the latter sets a stated maximum of imports, while the tariff sets a less precise maximum by levying an import tax so that less will be imported. In some cases quotas are parceled out to various countries (as for example in beef, where Argentina has a certain quota, Canada a certain quota, and so on); in others the quotas are general ones operated on a first-come first-served basis.

Quotas have long been used, but until recent years they were commonly

limited to special cases, such as the importation of narcotics. In the last decade, however, quotas have risen to major importance as a means of attaining the traditional tariff goals. Although the United States has not imposed quotas as widely as many other countries, cattle, milk and cream, fish, potatoes, fox furs, certain types of lumber, molasses and other sugar syrups, crude petroleum products, and miscellaneous other products are now subject to quota. Some countries have virtually their entire list of imports under such restrictions.

Quotas may be used as a means of granting special advantages to particular countries on particular commodities, though this is not the common use. Quotas, free of tariff, on specified commodities may be granted to other countries (usually in return for similar concessions), with only imports over this quota subject to the regular tariff. Thus the duty-free quota *may* be used as a device for permitting freer trade between particular nations. Generally, however, quotas are set up with the definite aim of limiting imports, and usually a tariff must be paid on the imports making up the quota.

In addition to formal quotas, governments often use other pretenses for establishing import quotas without using the name. In the United States this has been especially common in connection with food products, where strict examination is required for imports. To protect the public health, safety and morals, customs officials examine imports of animals and animal products, plants and plant products, and in general all foods, drugs, and cosmetics. Strict examination of such imports is of course desirable for the stated purpose of protecting the public health, safety and morals; but the examination has often been used as an informal means of excluding undesired imports, even though this exclusion was dictated by the motives behind the tariff rather than by the desire to protect the public health and safety. To what extent these examinations, here and in other countries, actually are used as informal quota devices cannot be accurately determined. It is certain that they are to some extent so used.

Another administrative method for shutting out imports is the exclusion of products violating standards of fair competition in international trade. For example, foreign products are required to bear marks showing the country of origin; thus they must not violate trademarks, patents, copyrights, and so on. Such requirements are necessary to protect domestic producers against unfair competition from abroad, but they may also be used as pretenses to exclude products beyond "fair competition" limits in order to achieve the results obtained by formal tariffs and quotas. Similarly, by making customs inspections very severe, imposing "red tape" on importers, overvaluing foreign imports so that they have to pay higher duties, and so on, customs officials can obtain the same sort of restriction on imports as is obtained by more formal methods. All such informal barriers to imports constitute an "invisible tariff," which may be very severe if a government desires to make it so through its administrative procedure in handling imports.

Export Subsidies

The major purpose of imposing tariffs and import quotas in depression periods is to stimulate production and increase domestic employment. By shutting out foreign products a larger share of the market is reserved for domestic producers. But it is also desirable to increase exports at the same time, since thereby production and employment in the export industries would be increased. However, unless some specific action is taken to the contrary, restricting a country's imports will curtail its exports, especially if retaliatory tariffs and quotas are imposed. Therefore, in order to stimulate exports in spite of import restrictions, special subsidies are sometimes given to exporters, making it profitable for them to produce and sell abroad even though decreased foreign demand lowers the price obtainable. When other countries have high tariff barriers, export subsidies permit domestic exporters to operate profitably, even though without the subsidy it would be impossible for them to produce for sale abroad. Export subsidies are simply public grants to producers of exportable goods to enable them to employ more workers through making it possible for them to sell abroad at prices lower than costs. The theory is that this will increase markets for domestic producers by stealing the foreign markets from foreign producers. The subsidy is given only to exporters, since if it were given to sellers in the domestic market this would simply be stealing the market from other domestic producers.

In agriculture the purpose of export subsidies has usually been somewhat different. Instead of trying to increase employment, the aim has been primarily to increase the income of farmers by raising prices of farm products. Most agricultural products are partially sold in the world market and their prices set in that market. In order to raise significantly the world price of cotton, for example, the United States would have to cut its output sharply. If we did so reduce output and raise price, the result would be to draw forth increased production from other countries. Therefore, in order to assure farmers higher prices for their products, export subsidies have been coupled with restrictions on the amount of crops permitted to flow into the domestic market. The domestic price is held up by this restriction, and farmers are given an export subsidy on the remainder of their output equal to the difference between the artificial domestic price and the lower world price so that they can sell the balance in the world market. This scheme gives the farmer higher prices at home and guarantees him against having to take a lower price on the amount held off the domestic market for sale abroad. This benefit for farmers, of course, comes at the expense of domestic consumers and taxpayers, and of foreign producers.

Among the more important commodities receiving export subsidies have been beet sugar, meat products, liquors, and iron and steel products, but the total list is a long one. The greatest project of this sort in the United States was the plan to subsidize farm exports during the 1920's,

though we have at times indirectly subsidized all exports through financing exporters by government funds at low interest rates.

Export subsidies, like tariffs and quotas, may temporarily increase domestic employment and benefit particular producer groups. Actually, however, it is almost certain that other nations will retaliate by raising their tariffs, putting quotas on imports, or possibly by giving competitive export subsidies to their own producers in order to permit them to steal American markets from our domestic producers. To the extent that retaliatory measures are taken by foreign governments, the effect of export subsidies is nullified, though their costs remain in both countries as burdens to domestic taxpayers or bondholders who furnish the funds for the subsidies. If set up to maintain a higher domestic price than the world market will support, their result is to increase price to the consumer and to maintain excess production.

Exchange Control

The use of tariffs, quotas, and export subsidies to increase domestic employment at the expense of foreign producers are direct stimuli to domestic employment. But since 1929 they have been increasingly supplemented by control of exchange rates to accomplish the same end. The case of England leaving the gold standard in 1931 is a prominent example. England went off gold partly because she wanted to be free to expand credit and check the price-income deflation in spite of gold outflows. But of correlative importance was the desire to increase British exports and thereby increase employment in England. The British expected that, as a result of going off gold, the pound would fall in value relative to other currencies since there would no longer be a specific gold content of the pound. Such a depreciation did in fact occur, to roughly \$4 to £1. It was then easier for Americans (foreigners in general) to buy British goods than before, since a pound could be obtained with less dollars than previously. Therefore, by going off gold and devaluing the pound relative to other currencies, the British obtained a stimulus to their exports, increasing employment and income in British export industries. Devaluation of the pound relative to other currencies was still another method of temporarily increasing British exports at the expense of other countries.

When England went off gold in order to obtain this export advantage, other nations began to retaliate lest they lose their markets to British producers. The gold standard was cast into discard by one country after another, and the world saw a great competitive devaluation of currencies. Until 1933-34 the dollar was maintained at its old gold level; dollar devaluation, when it came, was partly to gain freedom from automatic dependence on gold flows, partly to stimulate exports and give employment in export industries. With the formal United States devaluation proclamation of 1934, the franc, of all the world's major currencies, remained at its old gold level.

Although some temporary advantages were gained by this competitive

devaluation, especially at the expense of such countries as France, the main results were a great amount of uncertainty in international trade and lending because of the fluctuations in exchange rates and growing political enmity between nations. This currency warfare, of course, was superimposed on the higher tariffs, import quotas, export subsidies, and other less formal means of interference with international trade which were erecting cumulative import barriers in all countries.

As these competitive practices failed to bring many advantages for any nation, owing to widespread retaliation, some countries, especially in central Europe, resorted to still more drastic tactics. One of the first of these was "blocked exchanges," used by Germany and then by other countries, under which foreigners were prohibited from withdrawing payments received in Germany and forced instead to spend them in Germany, thus stimulating German employment. This control was increasingly linked in Germany with the rearmament program, under which the government exercised ever-tightening control over the entire economic system to expedite this major war aim. The government gradually took over all foreign exchange, limiting exchange usage to expenditures essential to rearmament and national welfare and removing all individual freedom in foreign payments. Simultaneously every effort was made to increase German exports to get more foreign exchange for purchasing needed materials abroad. Under these arrangements private individuals and firms were increasingly eliminated from foreign dealings, and foreign trade was more and more taken over by the government, especially in the German-dominated countries of central Europe. In many cases governments concluded "exchange clearing agreements." In others they went further yet and simply carried on large-scale barter; an example is the German barter of machinery for raw materials from several of the Balkan countries.

Where this barter stage was reached, the traditional mechanism of international trade had completely vanished. Even in less extreme cases, exchange rates have become of less and less significance. Furthermore, and perhaps of greatest importance, with increasing government control of international trade has necessarily come increasing government control of domestic economic life. If the government is to carry on trade by barter or other direct arrangement with other nations, it must control domestic production of the things it is trading. In the same way, it must control distribution of the goods received from other governments. Even with the less drastic step of exchange rationing, the government decides by the ways in which it rations exchange what imports to permit, and by controlling exchanges and giving export subsidies what exports to encourage. Increasingly the direction of economic activity has been removed from the price system and placed under the discretionary control of government officials. This centralization of control over economic activity has gone furthest in the dictator states, but even before the war, it had grown to major proportions in many more democratic nations.

Conclusion

If one takes an overall view of the whole wave of growing controls over international economic relations—tariffs, quotas, export subsidies, exchange controls, and others—three conclusions stand out. *First*, all such controls have contributed to the stifling of international trade which the world has suffered in the past decade, without gaining for any country very great advantages in the face of retaliatory action. *Second*, retaliatory measures have markedly increased international political friction and have undoubtedly helped precipitate the present world chaos. *Third*, with increasing government control over international trade necessarily has come increasing government control over domestic economic activity, with less and less power permitted to allocation of resources through the free price system and more and more control of economic life by discretionary government authorities.

Many writers who feel that individual and international economic freedom should be restored argue that there are three fundamental economic conditions that must be recognized as essential foundation-stones for re-establishment of a tolerably smooth-functioning and productive international economic order. *First* of these is a general downward revision of tariffs and other barriers in order to free international trade and lending, plus a reasonable guarantee against the renewal of competitive restrictions. Any given level of tariffs and other barriers is costly to all nations because of the inefficient allocation of resources that results. But the greatest losses and major disturbances come in the competitive restrictive struggles between nations such as have characterized the past decade. *Second* is the establishment of a reasonable degree of exchange stability and the elimination of competitive exchange depreciation. The advantages of relatively stable exchanges in expediting foreign trade and lending are substantial, but again far more important is the prevention of bitter and disruptive competitive depreciation struggles. *Third* is the provision of reasonably free access for all nations to the raw materials that they need. Although the importance of the lack of raw materials in causing political frictions and ultimately war has been overemphasized by many writers, both economic and political considerations point strongly toward giving to all countries free access to raw-material areas. This is only the logical counterpart of the general free trade argument; not only should nations be free to sell where they will but they should also be able to buy where they will.

Economists have long recognized the importance of these three considerations. Today the one great hope for their attainment appears to lie in a sweeping postwar economic reconstruction following an Allied victory. Except in times of unusual stress and change, vested national interests have consistently proved too strong to permit such comprehensive internationalism. But underlying these three requirements is another perhaps even more fundamental. As long as relatively full employment exists and depression is averted, countries may well refrain

from cut-throat international practices in the economic sphere. The prevention of unemployment, the establishment of a higher degree of economic security, the raising of planes of living—these are perhaps the greatest economic safeguards against international autocracy, though many other political and social pressures toward war and aggression would even then remain.

Some writers have argued that these ends on an international scale would require a single world currency, with monetary and fiscal policies throughout the world established in unison to prevent deflation. Others, however, suggest that were only a few of the major countries (possibly only England and the United States) to stabilize their domestic economies and adopt some sort of cooperative exchange arrangements to prevent serious exchange fluctuations, the adherence of other countries to this international bloc would go far toward keeping them all on an even economic keel. Still others have argued that what is required is an international economic supervisory body that would have the power to see that no nation adopted restrictive economic measures and that would advise or determine fiscal and monetary policies. This last proposal, of course, is closely related to the question whether an international political super-state or a revived and strengthened League of Nations is not a necessary condition for continuing international peace and order. Establishment of such an international economic control would necessarily mean the transfer to it of many of the powers of national economic sovereignty if the scheme were to function; and it is this fact that places perhaps insurmountable obstacles in the way of its achievement. But whatever the technical arrangement used, international like domestic economic order demands that individual nations "play the game according to the rules," rather than engage in competitive cut-throat restrictive practices to promote their own welfare at the expense of others. An orderly international economic system can function only on the basis of international peace and political "decency," and international political peace and cooperation can succeed only on the basis of an orderly international economic system. Economic friction generates political friction, and political friction generates economic friction. War is not caused by economic frictions alone, nor can peace be maintained by economic agreement alone, but economic forces do play a central role in the determination of international political relations.

Part XII

AGRICULTURE: A CASE STUDY

PREAMBLE

ECONOMICS deals ultimately with two kinds of practical questions: (1) What are likely to be the results of proposed measures? and (2) by what kinds of measures can desired results be obtained? If the long survey of economic analysis just concluded has any real usefulness, it is primarily in helping provide answers to these questions. It is proper, therefore, to test this framework of analytical procedures by applying it to a broad real-world problem—that of agriculture and its manifold difficulties—to see what general usefulness such techniques have in evaluating what has been done in recent years and in pointing toward what ought to be done in the future. In a single chapter, only broad questions may be raised and only general policies considered. Yet this is not altogether disadvantageous, for we are prone, in economics as elsewhere, to miss the forest for staring so hard at the individual trees.

CHAPTER 56

Agriculture: A Case Study

ALTHOUGH agricultural problems have arisen intermittently to plague politicians and economists throughout the history of the United States, it has only been in the decades since the First World War that "the farm problem" has become one of the great political and economic issues. Actually, "the farm problem" is a highly misleading term, for the troubles of the agricultural sector of our economy are varied and numerous. Agriculture is a complex of many "industries" producing a wide variety of products. Why, then, is there any more justification for speaking of "the farm problem" than of "the industrial problem" or "the commercial problem"? It is not entirely clear that there is, but certain characteristics common to most agricultural activities set this general sector of the economy apart sufficiently to justify its treatment as a separate case study.

The purpose of this chapter is not to set forth new principles disclosed by the study of agriculture but rather to apply to the problems of agriculture some of the relevant principles and methods of general economic analysis. There are few practical problems as well suited to this purpose as the mutually interrelated ones lumped under "the farm problem," for they involve the application of almost every principle of analysis economists possess. Output policies of the firm, price determination under competitive and noncompetitive conditions, the basic problem of allocation of resources to changing consumers' demand, monetary policy and cyclical business fluctuations, international trade—all these and others are directly involved in "the farm problem," and its solution, if indeed it has one.

Furthermore, "the farm problem" is of special interest because it shows so clearly the conflict between "economic logic" and political feasibility, a conflict that exists almost everywhere in our wide range of politico-economic problems but that comes more clearly to the surface here than almost anywhere else. Even though economic analysis may point directly to the need for some policy or to the unwisdom of some current action, the politician or administrator must face the realities of noneconomic pressures.

The Farmer's Economic Status in Society

The heart of "the farm problem" is the very low income received by the vast majority of farm families, both in absolute terms and relative to other

economic groups. In 1935-36, a typical "in between" boom and depression year before the sharp income increases of the war period, slightly over 600,000 farm families received direct relief, and the incomes of the other 6,000,000-odd farm families were distributed as shown in Table 56—1.¹

TABLE 56—1 *

NONRELIEF FARM FAMILIES BY INCOME LEVELS, 1935-36

<i>Income level (dollars)</i>	<i>Families (number)</i>	<i>Proportion of all farm families (per cent)</i>
Under 250	233,040	3.8
250 to 500	858,963	13.9
500 to 750	1,108,400	18.0
750 to 1,000	1,027,044	16.6
1,000 to 1,500	1,394,821	22.6
1,500 to 2,000	730,811	11.8
2,000 to 2,500	340,645	5.6
2,500 and over	473,834	7.7

* Data from *Consumer Incomes in the United States, 1935-36*.

Assuming that 85 per cent of the relief families received less than \$500 annual money income, these data indicate that approximately 1,600,000 families, or almost one quarter of all farm families received incomes of less than \$500, and that about 40 per cent of all farm families received less than \$750. In addition, it must be remembered that the typical farm family is larger than the 4-person group average for the entire economy. Indeed, the 25 per cent of the total population who live on farms are producing almost one third of all the children in the United States.

Aside from the question of money income received, farm families appear relatively worse off than urban groups by many other standards. For example, educational facilities and medical services are notably inferior in many rural areas. Although improvements are now being made rapidly, rural housing conditions are generally inferior to urban, both in buildings and equipment. For instance, even after the vigorous rural electrification program of the New Deal only 25 per cent of farm families have electricity for lighting, as compared with 95 per cent of urban homes. About 30 per cent of farm homes have water piped in the house as compared with 95 per cent of urban homes.

As against these facts, farm families have often somewhat better diets than comparable urban families, due partially to the use of fresh home-grown foods, more outdoor activity, and similar advantages, but these do not obscure the very low plane of living attained by vast numbers of farm families. With such boom conditions as have recently developed under war stimulus the farmers' absolute and relative position improves sharply,

¹ This table, and much of the other factual material in this section, is based on O. V. Wells, "Agriculture Today: An Appraisal of the Agricultural Problem," *Yearbook of Agriculture for 1940*.

but this fact alters little the long-run problems faced by farmers in the United States and by the policy makers who must deal with these problems.

A variety of causes for this income situation have been advanced—by farmers, by economists, by politicians, by everyone. As would be expected, from this variety of diagnoses have come a wide variety of proposals for “curing” the farm problem. As in the case of the proposal that wages paid on work relief projects be at trade-union levels, these proposals have centered around efforts to raise the prices received by farmers. Efforts have been made to hold up farm prices by buying up farm surpluses, by dumping surpluses abroad, and then by preventing the surpluses from being produced through crop reduction programs. Inevitably these efforts to control or fix farm prices have affected the allocation of productive resources, both in agriculture and elsewhere. The “income” and “resource allocation” problems have been intermingled, with results, so far at least, far from completely satisfactory.

In this chapter an endeavor is made to separate these two aspects of agricultural problems, first describing briefly the development of agricultural difficulties through the past few decades, then considering briefly the effects of some of the major farm programs, and lastly attempting to relate the “resource allocation” and “income” aspects of agricultural problems and programs.

Changing Fortunes in Agriculture

The production of most agricultural crops historically has more closely approximated conditions of pure competition than has any other major section of the economy. Since this has been the case, it is pertinent to ask, why then has not the long-run equilibrium adjustment described by formal competitive economic analysis come about? Why are farmers in particular in continued need of assistance? Why has it been considered necessary or desirable to take artificial action to hold up farm prices above what they otherwise would have been? The answers to these questions can be formulated only in the light of the historical position of agriculture relative to the rest of the economy.

Agriculture up to the 1920's

As in the case of most other newly developed countries, agriculture played the central part in the economic life of colonial and early nineteenth-century United States. With the development of industry and commerce on a major scale in the middle 1800's, the vast natural resources of the United States continued to give its farmers the basis for prosperous activities in the world market and in the rapidly expanding domestic market. In spite of temporary set-backs, these rapidly expanding domestic and world markets made the problems of agriculture almost entirely those of production. Technological improvements were sought, and found, to increase production and avert food shortages for the growing world population. Agriculture had its difficulties, to be sure. The open-

ing of the western prairies liquidated eastern grain farmers. Speculative land booms were repeatedly followed by sharp price declines, and there was a long "secular" depression in the late 1800's, which underlaid the Populist and Granger movements. But even though agriculture faced its difficulties and was increasing in scope at a less rapid rate than commerce and industry, it showed large absolute increases in persons employed and total production. The farmer was riding high on the wave of expansion.

The 1909-14 period is commonly looked upon as that period in recent times when agriculture was best off relative to the rest of the economy. It is this period that has been repeatedly chosen by farmers in recent years as the indicator of what would be a fair balance or "parity" position between agriculture and the rest of the economy in the formulation of farm legislation. But even in those years, there was already evidence of things to come. Already foreign production of agricultural products was booming and the comparative advantage of American farmers in many lines had slipped considerably, continuing a trend apparent well back in the nineteenth century. The foreign market for some types of meats had already been essentially lost, while foreign sales of grains and butter and cheese products had dropped sharply. At home, phenomenal increases in productivity based upon technological improvements were in some cases more than keeping pace with increasing demand and holding down farm prices. Yet on the whole, the picture seemed a bright one indeed to the farmer.

The First World War, though its immediate effects were highly disruptive, led to the greatest agricultural boom in the history of the United States, thrusting a great mountain of farm activity into the relatively declining trend of agriculture. As more and more European sources of foodstuffs were destroyed or shut off, the demand for American production steadily increased. The entrance of the United States into the war greatly intensified the need for agricultural products, especially in view of the widespread prosperity at home. Prices of farm products soared, and with them the prices of farm land. At the close of the war, instead of falling back or steadying, prices rose at an even more rapid rate as both the quantity and velocity of circulation of money and credit continued to rise. This speculative inflationary force centered especially in agriculture, buoyed up by the almost insatiable demand of a famished Europe. Wheat hit \$3, corn \$2, hogs \$19, cotton 40 cents; and prices of farm land soared accordingly. In 1910 the average value of land per farm in the United States was \$4,476. By 1920 this figure was \$8,503, and, to cite an extreme case, some of the best corn land in the Midwest reached the fabulous price of \$500 an acre, with almost all sales financed on mortgages at high interest rates. Huge debts were incurred to purchase farms at figures that could make profitable investments only if the unprecedently high prices for farm products continued to rise. Large acreages of previously submarginal land were drawn into production. Prevalent prices made it generally profitable to use extremely high-cost production methods, including use of high-cost areas.

Agriculture during the 'twenties

The unprecedentedly high postwar boom prices for farm products did not continue. The following year brought a precipitous crash in farm and industrial prices as the postwar speculative bubble collapsed. The Bureau of Labor Statistics wholesale price index fell from 157 in January, 1920, to 93 in June, 1921. The fall in farm product prices was even more disastrous. From a high of 170 in January, 1920, they crashed to 80 in June, 1921, a fall of well over 50 per cent in a year and a half. Agricultural land values, inflated by wildly speculative buying on borrowed money, could not be sustained and plummeted downward, leaving farmers stranded with huge mortgage debts at high interest rates on land now worth only a fraction of its mortgaged value. The accompanying business depression was sharp but short. Stimulated by the beginning of a great wave of construction activity and a quick return of optimism, business activity soon moved upward into the long prosperous period of the "roaring 'twenties." But agriculture found itself with still depressed prices and still more depressed land values, heavily burdened with mortgage debt. It was in this condition that agriculture entered the era of the prosperous 'twenties.

The widespread business prosperity of the 1920's was based on a combination of increased domestic and foreign demand, a general building boom, the introduction of numerous new products (especially the automobile) and new technological improvements in the production of old products, and a vast expansion of credit by the banks coupled with the introduction of large-scale installment selling. By contrast, agriculture was faced by a market that for the most part has ceased to grow and for some products was already contracting. In spite of soaring national and per capita incomes, people were not buying much more of many of the basic farm products, though there were great increases in the demand for such specialties as fruits and vegetables. Worse, the foreign market, on which the farmer had long depended for the disposition of the large portion of his production in many cases and some portion in almost all, had contracted sharply. Europe was in no position to buy except insofar as she could obtain loans from us to finance the purchases; and actually such loans were all that kept the American farmer from much direr straits during the 'twenties. Meanwhile, the rapid rate of population growth continued on farms, even though falling off considerably in urban areas, and technological improvements continued to increase the output per worker and per acre.

In such a situation, one would suppose that resources would shift from the relatively unprofitable agricultural pursuits into the booming industrial and commercial world, and that marginal and submarginal land would be withdrawn from production. To a considerable extent this was the case through the 'twenties, more so than most people probably realize. During this decade a net migration of over 6,000,000 people occurred from farms to urban areas, primarily industrial cities. Nearly 2 out of every

5 farm young people who reached their twentieth birthdays during the decade had moved to cities by 1930. Shifts of agricultural resources from one crop to another followed shifts in demand with remarkable closeness. To suppose that the economic forces of long-run adjustment to changed consumer demand were not at work would be a serious misconception, though it is a common one.

None the less, farming remained in a relatively depressed condition while industry and commerce boomed. What were the factors accounting for this slowness of adjustment? Why did not the lack of farm profits lead to a reduction of output sufficient to raise prices again to profitable levels? The experience of the war period had vividly demonstrated the quick, almost violent, upward adjustment of agricultural production to increased demand. But contraction in response to decreased demand has not been so easily accomplished.

Reasons for the Slow Downward Adjustment of Agricultural Production to Declining Demand

It is easy to overemphasize the particular difficulties involved in downward adjustments of agricultural production to decreases in demand. That downward adjustment is a painful and usually slow process for any industry is clearly shown by history. Almost never have producers, upon finding themselves faced with diminishing demand, been willing to admit the permanency of the situation. Almost never have they been willing to throw over their established businesses before going through a fairly long process of consuming their invested capital. Partially this is purely economic behavior, since it pays to continue production as long as income more than covers variable costs, even though losses are being incurred. Partially it is explained by the hope for better times in the future. Partially it is explained by the difficulties of finding attractive possibilities elsewhere for resources shifting out of declining industries. Similar problems have recently arisen, for example, in coal mining and in cotton manufacture. It would be a serious mistake to think that the difficulties of agriculture in this respect since the First World War, even though they have been on a very large scale, were unique.

Although the problem of downward agricultural production adjustments was not an unusual one, certain characteristics of agriculture are unique; moreover the difficulties there have been quantitatively so important as to merit special discussion. Although the particular circumstances vary considerably from one farm crop or animal product to the next, these differences may for the most part be ignored in order briefly to emphasize the major general factors at work.

In technical terms, the problem is largely one of very low elasticities of supply of the productive agents used in agriculture and fixed charges for them, coupled with a relatively inelastic demand for most farm products. Adjustment to decreased demand conditions, if it comes, must come either through withdrawal of productive agents or through lower returns to such agents (which returns are costs to the farmer-entrepreneur). If

productive agents are elastic in supply to an industry, they will be withdrawn when demand for the product falls and the price offered them declines. But if their elasticity of supply to the industry is low, their withdrawal in response to lowered demand will be less and slower.

In agriculture, a variety of factors combine to produce a highly inelastic supply for many agents. This fact alone would not be necessarily disastrous for the farmer were it not that in addition the returns to many of these relatively inelastic agents are often more or less fixed by previous contracts or basic living cost expenditures. The agents whose supply is usually very inelastic are primarily farm real estate and much farm labor (especially the farmer and his family), with the minimum return to each relatively fixed in most cases. Property taxes constitute another heavy relatively fixed cash outlay if tax foreclosure is to be avoided. As compared with such crucial fixed charges, variable costs ordinarily are of secondary quantitative importance, though expenditures for hired labor, seed, feed, and other such variable items usually have first claim on income. This situation of heavy fixed and low variable costs, coupled with the fact that withdrawal of agricultural resources from production often in fact involves complete cessation of production rather than merely withdrawal of variable agents, means that short period curtailments of production in response to decreased demand are seldom effective; and even over long periods downward adjustment often fails to occur except very painfully and slowly. A somewhat more detailed analysis of these reasons is worth while.

Real estate and associated fixed costs

The major investment involved in farming is the farm itself. Most farmers struggle to purchase a farm of their own at the earliest possible date, almost invariably financing the major portion of the purchase on borrowed funds and thereby incurring very heavy recurring fixed debt charges for principal and interest payments. The speculative boom in farm land during the First World War and the subsequent postwar period left most farmers with unprecedentedly heavy debt charges, with both principal and interest liabilities incurred at the boom period prices and interest rates. Heavy property taxes, based on inflated boom-year assessment values added to the burden of fixed cash outlays. Even had farm product prices remained at the high war-period levels these debt burdens would have been exceptionally heavy. In the face of the actual sharp price declines and continuing low farm product price levels through the 'twenties and the further precipitous declines in the early 'thirties, such fixed charges became virtually unbearable. But since variable costs were ordinarily so small a portion of total costs, it paid to continue production, even at heavy losses. Indeed, with such a heavy debt as many farmers faced, the lower the price of his products fell the more he was compelled to plan increased output in order to obtain the greatest possible amount of cash to meet his obligations. While such practices helped stave off creditors temporarily, mortgage foreclosures and to a

lesser extent tax foreclosures came in waves, first in the 'twenties and again in the early 'thirties.

The collapse of farm land values that accompanied the collapse in farm product prices intensified this difficulty, increasing the ratio of indebtedness to land values. Thus in 1920 the average debt per mortgaged farm in the United States was \$3,356, and by 1935, as a result of foreclosures and repayments, this had declined to \$2,899. But in spite of this decrease in average debt, the decline in land values was so much more severe that over this period the ratio of debt to land value rose from 29.1 per cent to 50.2 per cent.² This meant that not only were the prices of the farmer's products so low as to obviate the possibility of meeting fixed charges, but the associated decline in land values was so severe as to make it impossible for him to sell out at anything more than a fraction of the figure owed in many cases. If his debt was small enough to permit complete payment by sale of land at the depressed postwar figures, the farmer was often left with little or nothing to show for his large original cash payments and for his years of effort. If the debt was so large that interest charges and payments on principal were beyond the earning power of the farm and the land value was too low to permit full coverage of the debt, foreclosure by creditors was commonly the result.

The plight of the farmer in regard to the debt situation is plain. But the question remains to be answered as to why this process of foreclosure and liquidation did not more drastically reduce agricultural production. The answer lies largely in the fact that foreclosure of a farm mortgage ordinarily means only a shift in ownership rather than a withdrawal of the farm from production. Most farm land is economically shiftable between different crops, but during such a period the profitability of producing virtually all crops vanishes—there is little to be gained by shifting to another product. And except for shifting to alternative farm products there is virtually no alternative use to which most farm land can economically be transferred. The insurance company, bank, or other creditor that takes over the foreclosed farm finds itself with a heavy investment, and faces a depressed farm real estate market. In this situation, any return that can be obtained by rental or operation of the farm covering variable costs is likely to be better than permitting the property to lie idle or selling it on the weak market. In many cases the original owner remains on the farm as tenant for the foreclosing creditor. Farm real estate is a fixed agent that wears out much more slowly than most industrial and business buildings and equipment, and only marginal and submarginal land has any likelihood of being completely withdrawn from production on any substantial scale as a result of debt foreclosures. All available evidence shows clearly this very slow withdrawal of farm real estate from production even during these periods

² Figures from E. D. Johnson, "Agricultural Credit," *Yearbook of Agriculture for 1940*, p. 744.

of drastic liquidation, except as positive government steps were taken to force such withdrawal.

Taxes have already been mentioned as an additional heavy relatively fixed charge connected with farm real estate. The property tax is the primary farm tax, and it usually bulks large among cash farm costs. Many persons have long argued that farm land is overtaxed. Whatever the merits of this contention, it is undeniably true that assessment practices have been very bad and that assessed values have ordinarily lagged far behind actual value changes. In periods of declining prices, this means that the farmer remains burdened with heavy tax payments in the face of falling prices and land values, payments that cannot be avoided by production curtailment.

Depreciation of soil (usually called "depletion") and of buildings and equipment ordinarily constitute an important cost item, and much of this depreciation continues regardless of the level of production, representing another important relatively fixed cost that helps make it profitable to continue production even though total costs are not covered. Faced with a desperate situation, low income farmers have commonly failed to maintain the farms that they operate, either the soil itself or the buildings and equipment. Since other relatively fixed charges are very difficult to postpone in depression periods, maintenance of soil and equipment is often neglected except insofar as it is crucial to immediate continuance of operation. Indeed, the problems of soil misuse and wastage are tied up closely with the fact that farmers' incomes have often been so low and other claims on income so pressing that they have been unable or unwilling to make needed outlays for conservation and replacement.

In summary, it appears therefore that the inelasticity of the supply of farm real estate to farming as a related group of "industries," coupled with the highly fixed nature of many of the payments in connection with such real estate, make a downward adjustment of farm production a slow and painful process for the farmer; he is forced to bear a major share of the loss involved, either through negative profits or through lower returns on his investment, in case he owns his farm outright. Only through permitting depreciation to continue unattended or by obtaining concessions from creditors can he ordinarily gain temporary relief. Actually, of course, governmental action has been taken on a large scale, especially in the 'thirties, through debt moratoria and the provision of low-cost credit to ease these problems of farm adjustment in the debt field, and private concessions by creditors have been widespread. None the less, the fundamental difficulty of the farmer remained painfully apparent, and downward adjustment of production was retarded greatly by the factors discussed through both decades following the war.

Farm labor and associated fixed costs

As the low elasticity of supply of farm real estate to farming, coupled with associated fixed charges, underlay much of the resistance to produc-

tion curtailment in farming, so did the low elasticity of supply of farm labor and the relatively fixed nature of its charges play a similar part. Although heavy migration of farm population to urban areas did occur during prosperous periods, this flow virtually stopped and was even temporarily reversed in depression years when industrial and business employment became unavailable. Although farm labor is more mobile out of farming in response to downward pressure than is farm real estate, in times of stress this mobility becomes very low and the farm labor represents in most instances a very real fixed cost. Since the farmer utilizes so largely his own labor and that of his family, and since no higher income is clearly available elsewhere, he and his family become essentially fixed agents and their sustenance becomes virtually a fixed cost. Although the farmer's cash income may be almost entirely consumed by variable costs and only the most urgent of fixed costs, as long as he obtains enough to keep himself and his family going he is likely to cling to the farm.

Lack of alternative opportunities plays a major part in accounting for this low elasticity of supply of farm labor. The lack of occupational mobility among farm families is in considerable part directly traceable to the meager resources of many farm families and communities. In spite of substantial improvements in recent decades, educational facilities and methods in most rural communities lag far behind those in urban areas. Combined with this lack of educational opportunities, the low money income of most farm families fairly effectively bars farm youths from advanced training in lines requiring expensive education and training periods. Moreover, lack of alternative opportunities is much exaggerated by the restrictions on entry into many urban occupations. These are especially marked in some of the skilled trades, which might otherwise offer opportunities to the overflowing farm population. Finally the general depression of the 'thirties presented a gloomy picture to surplus farm population and urban unemployed alike in their attempts to find work opportunities.

Lack of information as to alternative opportunities, even when they exist, has been another major cause of immobility for some farm groups. This has been due in part to the failure of our governments until the present war to set up effective employment and information exchanges such as have been of major service in other countries. Although lack of this sorely needed action by the government has affected all groups, it has especially penalized farm youths in view of their special disadvantages and in view of the fact that each year piles up added "surplus population" on the farms unless a continuous flow into other sectors occurs.

Finally, experience has shown that a vast number of farmers prefer farming to other occupations to the extent of being willing to stay on the farm even though greater monetary returns are available elsewhere. This is especially true of persons of middle age or beyond. Of course this characteristic is not peculiar to farmers. Many persons do not

want to change their mode of life—coal miners do not want to become farmers any more than farmers want to become miners. These “non-economic” preferences are basic sociological facts, only especially apparent in farming. The quantitative importance of such “noneconomic” preferences for rural over urban life cannot be accurately measured, especially in view of the numerous other factors limiting the flow of agricultural resources into other sectors. But in many cases farmers do consider farming not only a means of livelihood but a mode of living to which they cling despite the most adverse circumstances, and to the extent that this preference is operative the flow of resources out of farming to more profitable alternatives is checked.

Production curtailment and inelastic farm demand

The failure of farmers to reduce production in accordance with falling market demand has proved especially disruptive because of the very low price elasticity of demand for most farm products. With such an inelastic demand, increased supply relative to the market demand leads to a very substantial fall in price and to a decrease in total farm income as a result of production in excess of the amount demanded at current prices. Thus inelasticity reinforces the more critical declines in demand itself that have occurred both in cycles and over longer periods. Therefore, the tendency of farm production to hold up in spite of decreasing demand in many lines in recent years has been especially apt to lead to price decreases and in many cases to a smaller total farm crop income than would have been obtained from smaller production.

Agricultural and General Business Fluctuations

That agricultural prosperity does not necessarily follow from general business prosperity is demonstrated by the relatively depressed state of most lines of agriculture during the 'twenties. Yet there is strong evidence that good and bad times in agriculture are closely dependent upon the level of national income and the state of general business activity. A prominent economist has emphasized this fact:

Agriculture is becoming more and more the football of business. Increasingly, industry is the chief buyer of agricultural products. When industry is prosperous, manufacture absorbs an immense amount of agricultural raw materials, and its wage earners, fully employed at good wages, buy large quantities of food products. When industry is depressed, the bottom drops out of the market for farm products.³

This does not mean that an exclusive causal relation runs from industrial to agricultural activity. Good or bad times in the two economic areas are closely interacting. Good business raises the demand for agricultural products, and higher income for farmers increases their demand for the products of industry. Agricultural fluctuations play an

³ Hansen, A. H.. “The Business Cycle and Its Relation to Agriculture,” *Journal of Farm Economics*, XIV (1932), p. 62.

important part in the general expansions and contractions that we know as booms and depressions. Some writers have even gone so far as to argue that fluctuations in agricultural production provide the major stimulating force for general business booms and depressions, but this thesis fails to withstand critical examination.

The highly competitive conditions in the selling market for farm products means that farm prices move rather flexibly in response to changing demand and supply conditions. In depression periods, such as that following 1929, farm prices fall much more sharply than do industrial prices, taken as a group. Following 1929, this was largely because agricultural production was essentially maintained through these years, while almost all lines of industrial production were sharply curtailed. The inelastic demand for most agricultural products made the price declines resulting from continued output in the face of falling demand especially serious.

Table 56—2 shows why it is so often said that agriculture took the great depression in lower prices while industry took it in reduced output with prices maintained. This statement, while it is a very loose one

TABLE 56—2*

INDUSTRIAL AND AGRICULTURAL PRODUCTION AND PRICE INDEXES, 1929-1939

	<i>Industrial production</i>	<i>Agricultural production</i>	<i>Industrial prices</i>	<i>Agricultural prices</i>
1929	100	100	100	100
1930	83	100	95	86
1931	68	106	81	60
1932	53	99	70	44
1933	63	96	71	48
1934	68	93	80	62
1935	79	91	82	74
1936	94	94	81	78
1937	103	108	85	83
1938	80	103	80	65
1939	93	106	79	64

* Reproduced from *Agriculture and the National Economy*, Temporary National Economic Committee Monograph No. 23, p. 39.

concealing many individual differences in the two areas, does emphasize that industry in general was able to restrict production, laying off workers, while agriculture found itself unable to follow this policy. The result was widespread unemployment and idle resources in industry, while in agriculture the burden of hard times was borne more directly by the entrepreneur owner-manager. This contrast was especially noticeable between agriculture and highly monopolized sectors of industry, and the difference has been used by many as a basis for the claim that agriculture should receive special governmental assistance.

The dependence of agriculture on general business activity becomes all the more crucial in the light of the population situation. Even during

the era of rapid agricultural expansion the high rural birth rate necessitated a constant flow of people from rural to urban areas if a "surplus" farm population was to be avoided. With the relative decline of agriculture in the economy during the last two decades, the necessity for this flow has become much more urgent. During the 'twenties, industry and commerce were able and eager to absorb these additional workers from the farms, and did so on a large scale. But with the depression of the 'thirties, industry and commerce not only were unable to absorb new rural workers but were even unable to provide work for urban residents. At the depth of the depression, there was even a net flow of population from the cities to rural areas, though little of this flow went to "commercial agriculture"—farms operated primarily for sale of basic products in the open market. Although this net return flow was soon reversed, the depressed level of industrial activity, coupled with monopolistic restraints by certain industrial producers and labor unions and the absence of effective employment and information exchanges, backed up this surplus production of youths on the farm. The loss of the market for its excess population was as great a blow to agriculture in the depression as the loss of the market for agricultural products.

With the revival of business activity in 1936-37 and again with the armament boom in 1940, agriculture found its own position much improved. However, by this time the effect of special governmental aid to farmers had become so widespread as to make very difficult a determination of the relative effects of each of the major causes of increased agricultural income.

Agriculture and International Trade

Just as the course of general business fluctuations has played a major part in determining the level of agricultural prosperity, so has the course of international trade played a crucial role in influencing farm welfare. The strong comparative advantage of American agriculture during the nineteenth century has already been noted, an advantage especially profitable because of the unprecedentedly rapid rate of population growth in both this country and Europe. And the great agricultural boom of the First World War period was based at least as much on the insatiable European demand for foodstuffs and materials as on the inflated domestic demand.

Following the war, a considerable portion of the foreign demand for American agricultural products vanished as war demands lapsed and production was resumed in the warring areas. Yet the great needs of a prostrate Europe during a period of reconstruction assured a continued high demand for American farm products if Europeans could find means of payment. This they managed to do largely as a result of the huge volume of foreign loans extended to Europe by private lenders and the government of the United States. All through the 'twenties the foreign market was supported in this fashion. In spite of this, increased production in other areas, notably South America, Russia, and the various

portions of the British Commonwealth, was providing increased competition for American farmers in the world market.

With the financial and economic collapse of 1929 and the almost complete cessation of United States loans to Europe, coupled with sharply increased United States tariffs under the Smoot-Hawley Act, American farmers found their foreign market reduced drastically, to a mere fraction of its previous size. This shrinkage continued through the depression, and even with the substantial revival under the Trade Agreements Program and better world business following 1936 our increased exports were more and more manufactured products in contrast with the earlier predominance of agricultural products.

Table 56—3 shows, more clearly than words can, the steady downward trend in the relative importance of the export demand for agriculture's basic products, and the especially sharp drop in the period since

TABLE 56—3 *

RELATIVE IMPORTANCE OF EXPORT DEMAND FOR SELECTED MAJOR FARM PRODUCTS

	Percentage contribution of exports to gross income from production of—			
	Cotton	Tobacco	Wheat	Pork and pork products
1874-78	70.2%	62.4%	28.1%	13.0%
1879-83	67.8	50.2	37.8	17.9
1884-88	67.2	50.1	29.3	12.4
1889-93	66.7	42.8	34.7	17.6
1894-98	69.2	41.6	34.3	18.8
1899-1903	67.3	39.8	34.5	20.9
1904-08	67.5	39.4	20.5	17.2
1909-13	67.5	41.4	18.0	13.3
1914-18	48.8	37.7	34.4	18.8
1919-23	58.4	44.0	34.9	21.1
1924-28	58.7	43.0	27.0	12.4
1929-33	56.4	38.8	15.7	6.9
1934-37	43.6	36.7	7.6	2.7

* Reproduced from *Agriculture and the National Economy*, Temporary National Economic Committee Monograph No. 23, p. 6.

1929. The figures in the table stop after 1937, to avoid the special influence of the onset of the Second World War. These data reflect not only the declining comparative advantage of American agriculture in regard to these basic products, but also the steeply rising tariffs and other trade barriers, both here and abroad. In part the table also mirrors, of course, the growing domestic demand that in some cases absorbed an increasing part of domestic production. But the farmer has surely been one of the greatest sufferers from the stifling of foreign trade attendant upon the increasing economic "nationalism" of recent years.

Whether the postwar export market for American agricultural products

will approach pre-1930 levels depends in considerable part on the outcome of the war. If the Allies emerge victorious and a world of relatively free and cooperative economic relations is established, we may certainly expect a considerable revival of our export markets, assuming that we participate by lowering our import barriers. But even in that case, there appears little evidence that the American farmer can ever again, except under special circumstances such as postwar reconstruction, count on the foreign market that he knew during the period before the First World War or even during the 'twenties. Competitive producing areas for many of our basic agricultural products have become well established, and our comparative advantage appears to have shifted strongly in the direction of manufactured products, particularly the mass production goods of metals and machinery.

Agriculture's Claim to Special Aid

Although their demands for special aid have unfortunately often been based on less defensible arguments, farmers may reasonably lay claim on a number of grounds to governmental assistance in meeting their manifold problems. (1) As a low income group, generally speaking, they may ask special aid to establish minimum planes of living. (2) In view of the large farm families that have provided more than their share of the population, they may ask special remuneration. Neither of these claims applies only to farming. Both are based on special circumstances under which nonfarming groups may also claim assistance. (3) The farmer sees many other groups in society receiving special privileges and special advantages, and may demand that if these are to continue he too should receive special aid. (4) Agricultural prices, incomes, and production conditions are particularly erratic, subject to especially violent fluctuations, and the farmer may seek aid in meeting this problem. (5) If, as is sometimes argued, agriculture as a way of life is particularly fundamental to our individualistic democracy and provides a considerable degree of social stability, agriculture may claim special subsidies to enable it better to carry out this socio-political function.

1. The income status of farmers. The very low incomes received by many farmers relative to other economic groups are frequently presented as an argument for special assistance to agriculture. The farmer's case for special aid here rests not on his being a farmer, but on his being a human being with a very low income. If we believe in some sort of minimum plane of living for all and in less inequality of income, a large portion of farm families are among those meriting special assistance. It is important to note, however, that this is fundamentally a "personal" income problem rather than an "industry" problem. Although many farmers are in the lowest income groups, many are not, and there is little justification on the "income" score for helping the high income along with the low income farmers. Indeed, many economists argue that fundamental weaknesses of our recent farm-aid programs stem from the

common failure to distinguish between the personal income-distribution problem and the "industry" claims of agriculture for special assistance.

2. Large farm families. The second ground on which farmers may claim special governmental aid is the fact that there is no provision in the price system for paying them for the expenditures of money and effort that they make in bearing, rearing, and educating a large number of children who later move into other economic areas. Economically, human beings are capital resources of the nation as much as are buildings and machinery, and it may therefore be argued that the nation should remunerate the rural population for the contribution it makes to the national welfare in this connection. This again is not peculiarly a farm problem; large families are found in urban areas as well, but agriculture has long produced a "surplus" population. Many persons feel that agricultural youths are likely to become especially desirable workers and citizens; but if the primary intent is simply subsidizing large families in general, large urban families might equally well be included in the program.

3. Special aid to offset artificial advantages of other groups. The third ground on which farmers may claim special aid is that they are injured by special privileges granted other groups, and hence that they deserve offsetting special assistance. The high American tariff on industrial products has long forced the farmer to pay high domestic prices while shutting off his own foreign markets. Monopolistic conditions have been blinked at by governmental officials in a vast number of industrial products, in spite of laws purporting to outlaw monopoly in restraint of trade. Labor unions in many cases have been permitted to exercise monopolistic restraints to raise wages, raising the prices paid by farmers and restricting movement from the farm to urban occupations. In the face of these facts, the farmer long had to sell in an essentially competitive market, with very little in the way of special advantages.

Many economists object to this argument for special aid to agriculture, maintaining that the answer lies in the removal of special privileges for other groups rather than in building up still more special privileges and restrictions. They point to the very real danger of an ever-increasing spiral of special privileges and restrictions for each group to offset advantages of the other. Correct though this analysis may be, the farmer understandably feels that he has a claim to assistance if special privileges for other groups continue.

4. Erratic agricultural prices and incomes. There is probably no more unpredictable sphere of economic activity than farming. Farm prices, costs, and incomes sweep up and down with a frequency and violence that makes commercial agriculture risky in the extreme. Demands are relatively very inelastic particularly for some of the basic grain crops. Industrial demands for farm raw materials, such as cotton, are peculiarly sensitive to fluctuations in general business activity. Changes in international politics play havoc with foreign markets for

major crops. High fixed costs in agriculture make supply as a whole highly inelastic although adjustments between crops are quite easily accomplished. Even such adjustments, however, can be made only once a year. Finally, planning output in agriculture, whether by individuals or on a group basis, is not like planning output in industry. The vagaries of nature may upset the best-laid plans. Droughts have been the one sure method of major crop reductions that have brought major price increases. But good weather conditions have been equally effective in producing bumper crops that have flooded the market and depressed prices. Carefully as the individual farmer may plan, he is never sure what his output will be; and carefully as the agricultural industry may attempt to adjust production to expected demand, there is never any certainty that the planned adjustment will actually occur. The highly competitive position of the farmer plus the natural hazards of his occupation place him in a very insecure position relative to most other entrepreneurs.

There is little doubt that the uncertainty in commercial farming presents a problem in the solution of which the farmer has a claim to special governmental assistance. This claim rests on benefits to the economy as a whole as well as to the farmer. If many of the uncertainties of production and price can be eliminated by governmental assistance, much more efficient use of agricultural resources will be assured. If the flow of farm commodities into the market could be stabilized, consumers and industry as well as farmers would benefit greatly.

5. The agricultural way of life. Many people feel that farming as a way of life involves more community ties and greater social stability than is characteristic of other occupations. It is argued that community living and stable personal and social relations may be an anchor in a stormy social scene, where violent social and political crises and increasing strains on individuals might otherwise result in upheavals even more drastic. The significance of a vital agriculture as a central core in the functioning of individualist democracy is argued by many sociologists and political scientists. On the other hand, the drift of agriculture into more commercialized farming modifies these characteristics of the agricultural way of life as it has been known in the past, and in fact the major features of the agricultural program have focused on the "least agricultural" aspects of agriculture, that is the commercialized sectors of agriculture. Moreover, dilemmas arise in the conflicts between this goal and maximizing the plane of living of the entire population. The validity of the argument that agriculture should be encouraged as a way of life depends on many things consideration of which is in the main beyond the scope of this study. There is no "economic" basis for such an argument, but this by no means proves that the argument is therefore invalid.

6. Other claims. Valid or partially valid though these five claims to special assistance may be, they surely do not warrant a blanket justification for special aid to the farmer every time he finds himself in trouble

for whatever reason. If the farmer speculates unwisely in land and finds himself "left out on a limb," he has neither more nor less claim to sympathy than any other speculator. If he goes "overboard" for the production of grapefruit when consumers prefer oranges, he has no more claim to special assistance than would an industrialist who had guessed wrong on consumer demand. And most of all, if consumer demand shifts away from farm products more and more to nonfarm products, the farmer has economically no more claim to permanent subsidy to induce him to produce in excess of the level indicated by consumer demand than does the buggy manufacturer or the producer of women's high-top shoes. Unfortunately, it is this last and most fundamental factor of adjustment to shifting consumer demand that is often lost sight of in the midst of more spectacular arguments advanced by special interest groups.

Farm Aid—1921–1933

That our recent experiments in special aid to agriculture have helped agriculture, that they have failed to solve "the farm problem," that they have prevented the downward adjustment of agricultural output to decreasing demand, that they have speeded recovery, that they have impeded recovery—all these and many other claims have been bruited about in recent years. This section and the next are devoted to a brief survey of some of the major farm-aid programs undertaken recently, to the historical background of the problems that led to these programs, and to brief comments on the justification for special aid to agriculture.

As early as 1921 farmers began their organized pressure for special aid to agriculture, and Congress created a special Joint Commission of Agricultural Inquiry to investigate the plight of the farmer. This pressure for special assistance grew through the 'twenties as agriculture remained relatively depressed, and various legislative steps were taken to help farmers, notably the establishment of the Federal Intermediate Credit Banks, the law exempting farm cooperatives from provisions of the anti-trust laws, and a higher tariff on wheat, beef, and cattle. By 1927, enough strength had been mustered in Congress to pass the so-called McNary-Haugen Bill, providing for stabilization of the prices of certain basic farm products at "profitable" levels, but this legislation was vetoed by President Coolidge. The following year a similar, but somewhat broader, act was passed, and was again vetoed. Finally, in 1929 the Agricultural Marketing Act was passed and signed by President Hoover, setting up the Federal Farm Board with the stated purpose of encouraging "orderly marketing" of farm products. The pressure for "orderly marketing" legislation had originally been aimed at smoothing out seasonal variations, but gradually had been transformed into a broad program of continued widespread support of agricultural prices, reflecting the growing power and demands of agricultural pressure groups. For this purpose the Farm Board was given \$500,000,000, a sum which at that time appeared very large. In the same year, the Smoot-Hawley tariff provided higher

duties on farm products imported, along with major increases on industrial products. As so many times previously, economic policy developed largely as a bargain between the "farm West" and the "industrial East." Failing in its earlier demands, the farm bloc finally compromised on higher industrial tariffs in exchange for the Farm Board provisions.

By the summer of 1929 it was apparent that the prices of basic farm products, especially wheat and cotton, were going to fall because of bumper crops, large carry-overs, and declining European markets, now heavily invaded by Russian, Argentine, and Canadian wheat and by cotton from various new areas. The Farm Board therefore undertook stabilization operations in wheat, making loans to farmers' cooperatives to buy up and hold wheat off the market to avoid further depressing prices. By July 1, 1929, it was estimated that 614,000,000 bushels of wheat were in storage throughout the world, as compared with more or less "normal" holdings of about 275,000,000 bushels in 1926. The precipitous general economic decline in the fall of 1929 and during 1930, coupled with the almost complete cessation of American loans to Europe, led to a sharp break in wheat prices. The Board faced the painful prospect of bumper crops, tremendous carry-over, vanishing demand, and dwindling resources. Into the gap it plunged, undertaking to buy up wheat held by cooperatives and some newly produced wheat, but excluding wheat held by other persons or groups. By April 1931 the Board's holdings of wheat had increased to 200,000,000 bushels, the price of wheat on the open market had fallen drastically, and the Board's resources were practically exhausted.

The stabilization operations in cotton were similar in nature. Through loans to cooperatives the Farm Board attempted to hold surplus supplies off the market and so to hold up the price, but the efforts of the Board were of little avail against the basic forces of deepening depression, huge carry-overs, and bumper crops. By 1931 the Board found itself with huge supplies of cotton on hand without having more than temporarily checked the downward plunge.

That the Board's efforts to maintain farm prices and incomes by buying up surpluses had proved largely ineffective was generally admitted. Already growing support was to be found for the idea of preventing the emergence of surpluses through production control rather than by buying them up after production. It was this new attempt to raise farm prices and incomes by preventing "overproduction" that provided the core of the new and drastic Agricultural Adjustment Administration program, enacted as one of the first major pieces of New Deal legislation in May, 1933.

Farm Aid Programs Since 1933

The positive program of the New Deal to aid the farmer has passed through several stages since 1933. In spite of the diversity of the measures enacted, and the changes from year to year in those that survived, it is possible to distinguish seven major steps: (1) The first, basic to

the entire program, has been the effort to control production by limiting acreage through various expedients. This production control program has been supplemented by, and, in the case of certain basic crops, virtually identified with a system of marketing quotas. For specialty crops and milk and dairy products, producers' marketing agreements have been used instead of marketing quotas. (2) The second, closely following the old Farm Board, has been a program of crop loans and storage provisions intended to hold surplus crops off the market. (3) The third has been a series of measures intended to speed disposal of surplus crops, notably the food and cotton stamp plans of the Federal Surplus Commodities Corporation. (4) Attempts to restore the foreign market through the Trade Agreements Program and export subsidies constituted the fourth. (5) The fifth has been the effort to ease the debt and interest burdens of the farmer through provision of low-cost farm credit. (6) Sixth has been direct aid to supplement farm income, including the "parity income" payments to participating farmers and the activities of the Farm Security Administration to aid low income farm families. (7) Seventh have been direct efforts to promote the more "economic" or "scientific" use of resources in farming, including the activities of the Federal Crop Insurance Corporation intended to remove some of the uncertainties of natural hazards to production, and the broad, long-established program of scientific research, reports, and guidance for farmers through the Department of Agriculture, the Agricultural Experiment Stations, the county-agent system, and so on. And important throughout the entire program, though especially prominent in (1) and (7), have been efforts to facilitate conservation of our national resources.

Underlying much of this program are the concepts of "parity prices" and "parity income" for farmers, developed by the farm pressure groups during the 'twenties and continued with growing force through the 'thirties. The gist of these concepts is that agriculture should be as well off relative to the rest of the economy as it was during the "base period" of 1909-14, which is that period in recent times in which agricultural prices were generally highest relative to other prices. For practical purposes the concept of "parity prices" has been the important one, with farm legislation framed to raise individual farm prices to that level, relative to nonfarm prices, that they occupied in the base period. Essentially, the "parity" argument amounts to the proposition that, whatever the dictates of shifting consumer demand and changing technological conditions, artificial action should be taken to raise and support agricultural prices and income at their exceptionally favorable 1909-14 levels.⁴ Although the parity concept is fundamentally only a rule-of-thumb concept for farm pressure groups, it has been very effectively used by them on repeated occasions, and it has often obscured their basic demands for

⁴ Except, among the major crops, for tobacco, which has been given a more recent base period in view of the fact that 1909-14 was relatively less favorable for it than the 'twenties with their greatly increased demand for cigarettes.

subsidies under what had the appearance of being a "scientific" formula.⁵ Apparent as the fallacy of the parity concept must be, it has been accorded a tremendous amount of support. "Parity" is a good rallying slogan. Of course opponents of the "parity" concept and the farm measures enacted have not been silent. Why, asked the critics, would it not be equally wise to subsidize buggy-makers to the extent of maintaining prices and incomes in that industry at their 1905-06 levels since those appear to have been exceptionally good years for the buggy trade? But still the program has prospered in Congress. Any broad case study in economics must recognize the importance of political pressures and compromises.

That the program undertaken to aid agriculture has been comprehensive is apparent. Indeed, the joking description of the work of the major agricultural agencies current in Washington during the period—that AAA paid the farmers not to raise the crops, Commodity Credit Corporation paid them (through loans) to store the surplus AAA did not prevent, and Federal Surplus Commodities Corporation bought up and gave away what surplus was left over—though only a half truth, has an uncomfortably penetrating ring. The program has broad ramifications.

Production control and marketing quotas under AAA

The core of the New Deal farm program, ever since its inauguration in 1933, has been the effort to prevent crop "surpluses" and hold up prices by restricting production. The primary method used has been limitation of acreage planted, supplemented by a system of marketing quotas in cotton, tobacco, wheat, and rice. The arrangement thus is strikingly similar to the restrictive methods of private industrial oligopolists whereby total output is limited and the market "divided up" with price maintained by agreement.

The production control rationale of the program was plainly stated in the original Agricultural Adjustment Act of 1933. Payments were made to all farmers who restricted acreage in accordance with the allotments made by AAA administrators, the payments being financed by a processing tax on the commodities involved. No legal compulsion was imposed on farmers to cooperate in the program, but the large benefit payments, coupled with the social opprobrium attendant on noncooperation, brought a high percentage of participation. National allotments for each crop were determined in accordance with the law by national administrators, and apportionments were made to each state, which in turn apportioned its acreage allotment among counties and communities through local administrative committees composed of the farmers themselves. Acreage allotments were based largely on the proportion produced by each area in preceding years but with adjustments for type of farming, crop-rotation

⁵ For a remarkable, straight-forward recognition that the farm program has been at bottom simply a mass subsidy, see the valedictory statement of Henry A. Wallace as Secretary of Agriculture in the *Annual Report of the Secretary of Agriculture for 1940*, p. 184.

practices, type of labor and equipment available for production, and so on. The dividing of allotments among individual farmers has been the cause of much resentment on the part of those who felt they deserved larger quotas. In part this was due to inadequate and inaccurate data used by the administrators (later improved), in part to changes in individual farming operations more rapid than provided for in the formulation of the program, and in part simply to the natural desire of each farmer to benefit individually from artificially supported prices by obtaining the largest possible quota.

The AAA was declared unconstitutional in 1936, on the grounds that taxing one group (processors) to subsidize another was unconstitutional, as class legislation; but a new act was immediately passed eliminating the processing-tax feature and being strengthened by further relating production control to a program for conservation of natural resources, a phase of the original AAA that had been comparatively little stressed. However, production control under this act was highly ineffective, and in 1938 a broad new program of control was enacted, still in large part nominally for the purpose of soil conservation. Although the Supreme Court had become more sympathetic to such legislation, production control sounds much better under the name of soil conservation; besides, the widespread public resentment at the killing of the little pigs and plowing under of every third row of cotton in 1933 in the face of great human need had not been forgotten.

Under the 1938 Act, acreage allotments were again determined in roughly the same manner as under the 1933 version, with the national allotment for each basic crop determined so as to provide for "normal" needs plus an adequate reserve supply in order to assure an "Ever-Normal Granary" in the event of years of shortage. Included were cotton, corn, wheat, rice, tobacco, potatoes, and peanuts. An allotment was determined for all other "soil-depleting" crops as a group. Acceptance of the acreage allotments was again voluntary, with benefit payments only to participants. However, to ensure that nonparticipants would not reap the benefits of reduction by participants, marketing quotas were also permitted to be established where necessary. These marketing quotas, which limit the amount to be marketed by each producer (usually identical with that envisaged by his acreage allotment), may be proposed by the Secretary of Agriculture, and are voted upon by producers of the crop involved. If two thirds of those voting favor establishment of the quotas, they become compulsory for all producers, and noncompliance is punishable by fines. Enforcement of such quotas is practical only for crops that necessarily flow through established processing channels, and they have been applied mainly to cotton, tobacco, rice, and wheat. Even there, substantial numbers of producers have failed to stay within their quotas, preferring to pay the penalties imposed in order to produce and sell more at the higher prices maintained. However, in view of the failure of acreage control satisfactorily to restrict production, marketing quotas have come more and more to play the central role in the produc-

tion control program. For milk and dairy products and for specialty crops (vegetables, fruits, nuts, and so on) marketing agreements are encouraged in place of formal market quotas. Under the marketing agreements program, producers meet together and determine how and how much of their product should be marketed to raise their incomes, and, if he approves, the Secretary of Agriculture promulgates the marketing agreement, which then becomes binding upon all producers in the industry. The striking similarity of these production and marketing control schemes to the cartel or "share-the-market" arrangements characterizing many industrial groups is obvious.

How effective have these production control programs proved? The evidence available indicates that the acreage allotment programs, unsupplemented by marketing quotas, have proved an ineffective device for controlling production. Substantial groups of producers have remained outside the program, and total output even of participants has increased in many cases. In corn, for example, during the 1937-39 period acreage planted in the six central corn belt states was reduced 8 per cent from the 1929-32 period, but corn production in these states actually averaged 17 per cent higher. This was in part accomplished by increased yield per acre resulting from use of hybrid seed and more efficient productive methods stimulated by the higher price and restricted acreage, and in part by weather conditions. Moreover, the corn acreage restriction induced a widespread expansion in substitute crops, especially soybeans and the more productive legumes, so that the total feed supply of such crops was increased by much more than the increase in corn production. The possibilities for such increases were more limited in most other crops, but the conclusion must stand generally that acreage control has simply not proved effective in reducing production. Everywhere, increased yields per acre stimulated by higher prices have in considerable part, or completely, offset acreage reductions. Naturally the restriction in those crops covered by marketing quotas in addition has been more effective, but even there the results have been far from those desired by AAA planners.

With United States' entrance into the war, the AAA³ program has finally been readjusted to increase substantially production of specially needed products, such as peanuts and soybeans, on land withdrawn to conform to quotas on "soil-destroying" crops, although price maintenance at artificial levels for more plentiful farm products has continued a major aim. Presumably, increasing emphasis will be placed on increasing output in particular lines as shortages even in farm products pinch more and more, especially in connection with foreign needs.

Crop loans and storage

In the 1938 Act, crop loans were restored to a position of prominence as a part of the "Ever-Normal Granary" program. The purposes underlying the provision for high commodity loans varied for different crops, but the fundamental one was to assist production control in raising and

maintaining an "ever-normal granary"—a reserve supply sufficient to smooth out the ups and downs of annual production.

The manner in which these loans are made is a complicated one. The Commodity Credit Corporation is instructed to make loans at some per cent of the "parity price" of the commodity in question—from 52 to 75 under the 1938 Act; 85 under the 1941 Act; and up to over 100 under the 1942 wartime provisions, which in effect set price floors at or well above parity levels for most basic farm products. Parity price for the crop is that price at which it is in the same relation to nonfarm prices as during the base period. This base period has been 1909–14 for all major crops except tobacco, whose base period was changed to 1919–29 for some types and 1934–38 for others, the periods in which tobacco was in an exceptionally favorable relative price position. Once parity price for a crop is determined, loans are made at the rates specified upon security of the crops stored, except that nonparticipants in the production control programs can obtain loans only under certain restrictions and to a lower per cent of parity price than specified for participants. Loans may be made on any agricultural commodity, including dairy products.

Unlike other secured loans, loans made by the CCC are specifically without recourse. That is, if a 75 cent loan is made on corn and the price falls to 60 cents, the farmer may simply turn the stored corn over to the CCC and leave it to bear the loss. The CCC has no claim against him for the additional 15 cents per bushel. Thus, in effect, the crop-loan program amounts to a promise to buy the crop at the loan level in the event that the price goes lower and the farmer wishes to unload, while permitting the farmer to pay off his loan and sell the crop in the event that the price rises sufficiently to make it profitable to do so. From the farmer's point of view the scheme amounts to realization of the old saying, "Heads I win; tails you [the government] lose."

The crop loan and storage program early accumulated surpluses in storage that many economists considered ample ever-normal granary supplies; and by summer 1941, in spite of the booming demand for cotton which liquidated many loans, CCC had almost $1\frac{1}{4}$ billion dollars in purchases on hand and loans outstanding, secured by vast supplies of such basic crops as cotton, wheat, corn, and tobacco. Up to the great demands of wartime production on raw materials and the increased demand for food-stuffs with returning prosperity, the conclusion was hard to escape that the crop loans made had been seriously out of line with the prices indicated by economic demand and supply forces. None the less, the huge commodity stocks have helped to ease war shortages in those lines to the extent that they have been released by CCC for use. But even in the midst of war shortages and inflation, the "farm bloc" fought release of accumulated stocks lest the rise in farm prices be "unfairly" restrained.

The surplus disposal program

The government has endeavored to combine the aims of aiding low income families and disposing of accumulated farm surpluses through

the relief distribution and food-stamp plans of the Federal Surplus Commodities Corporation. Upon instruction by the Secretary of Agriculture, FSCC bought up in the market crops whose prices were severely depressed and distributed these crops free to relief clients through state and local relief agencies. For the most part, FSCC has dealt only in specialty crops and dairy products suitable for direct consumption, though in 1940 a cotton disposal program was instituted on a small scale. More recently, the food-stamp plan was inaugurated, whereby FSCC sold orange colored "food stamps" to low income families and with every \$1 orange stamp gave a 50-cent blue stamp. The orange stamps could be used by the purchaser at grocery stores in the purchase of any commodity, the blue only for those commodities of which the FSCC had declared surpluses existed. The grocer then cashed the stamps for full value from the FSCC. The virtues of this plan are claimed to be that it released the surpluses through regular market channels, improving the business of the participating retailers and giving the low income families more freedom of choice than they would have had through direct commodity distribution. The primary offsetting disadvantage appears to be that the food stamp plan was somewhat more expensive than direct relief distribution of the commodities.

During the four years 1935-39 nearly three billion pounds of surplus foods were distributed under the direct relief arrangement. The food-stamp plan was inaugurated in some cities in early 1939, and spread rapidly. However, it was never granted sufficient funds by Congress to assume a major role in the agricultural and relief programs, and did not become general throughout the country. With the advent of war shortages it was quickly discarded.

The trade agreements program and export subsidies

Narrowing international markets, the general resistance to outright tariff reductions, and the importance of opening closing doors to American exports led to the development of the Reciprocal Trade Agreements Program in 1934. Previous tariffs had been extremely high and this seemed a reasonable and politically feasible way of paring them down and expanding international trade. Agriculture benefited from this program in two ways. By lowering our tariffs on imported goods of all sorts, we indirectly gave other countries increased purchasing power over our agricultural (and other) exports. More directly, many of the agreements improved the opportunities to sell farm products abroad by lowering foreign tariffs and quotas that had been erected against them.

Simultaneously with the Trade Agreements Program, farmers were continuing their vociferous claims for export subsidies on farm products, to permit them to dump surpluses in other countries without incurring losses while helping to maintain prices within the United States, a practice already discussed. Obviously, this program was directly in conflict with the aim of the Trade Agreements Program toward freer trade and a good neighbor policy. Not only do such export subsidies conflict with the freer

trade ideal, but they invite retaliation both politically and economically. None the less, from a short-run point of view they appealed strongly to the interests of certain farm groups. And while the Department of State labored to promote trade agreements through the late 'thirties, the Department of Agriculture poured large sums into export subsidies, primarily on wheat and cotton, even though this phase of farm aid has received virtually no publicity. To the economist, the choice between the two methods of stimulating exports is not a difficult one.

The farm credit program

The major importance of interest and debt burdens in the difficulties of the farmer has already been indicated. Early in the depression of the 1930's farmers in all parts of the country were in dire straits in regard to debt charges, and the practice of granting state-wide and local debt moratoria spread rapidly, stimulated by farmers' "strikes" in some areas. Beginning in 1933 and continuing through 1935, an extensive program of refinancing farm mortgages at lower interest rates, totaling almost \$2,000,000,000, was carried out by the Federal Land Banks and the Land Bank Commissioner. A complex structure of credit agencies designed to provide credit for farmers has rapidly grown up, the nature and scope of which has been indicated in a previous chapter⁶ and need not be restated here.

The extent to which this system of federal credit agencies operates to provide a subsidy for farmers is not clear. In many cases, federal credit has filled a need of farmers not met by other sources of funds except at exorbitant rates, especially on loans of intermediate length—about one to five years. On the other hand, private lenders have argued tellingly that they are unable to compete on equal terms with government agencies that meet expenses from federal funds. There can be no doubt that the government credit program has drastically reduced the rates on farm loans of almost all classes. Whether this has come as the result of newly introduced competition in a monopolistic field or as a result of government subsidy must wait for a definitive answer. Probably both explanations are in part correct.

Additional direct aid to farmers

In addition to the various steps to raise and maintain farm prices and to make credit available to the farmer, several other types of direct special aid have been enacted.

The Agricultural Adjustment Acts have carried provisions for "parity income payments" in addition to the provisions already noted. These provisions have been stated on a basis of "if and when appropriations are made by Congress." When Congress has seen fit to appropriate special funds, these funds are parceled out by the AAA to farmers producing the basic commodities in proportion as the prices of these commodities are

⁶ "The Banks and the Quantity of Credit."

below their parity prices. This provision has been intended as an additional boost to farm income in light of the emphasis on "parity income" and of the fact that, until 1941, most farm prices had never closely approached "parity" in spite of the vast farm programs. For the most part parity income payments have not been part of the administration's farm program, but instead have been tacked on by Congress. The \$212,000,000 appropriation of 1938 was the largest made for this purpose.

Aid under all the programs so far mentioned, with the exception of farm credit, has gone to producers on the basis of their reduction of production or, through higher prices and loans, on the basis of the amount produced. By contrast, the Farm Security Administration was set up for the special purpose of extending aid to low income and underprivileged farm families. Activities of FSA have included a program of rehabilitation for low income families, emergency loans and grants to farmers who are unable to obtain aid or credit through regular channels, farm-debt adjustments, special assistance to tenants attempting to purchase farms, organization of medical and community service cooperatives, provisions of camps for migrant farm workers, and other such programs.

In 1939 FSA extended aid to more than 600,000 needy farm families in planning and financing farm rehabilitation. This aid included not only financial assistance but careful and complete planning of the future program of farm operation and improvement, both in the field and in the home. In all activities, assistance has been made in the form of loans where feasible, but through outright gifts in cases where the straits of the farmer were exceptionally dire. FSA assistance has proved especially valuable to the low income farm belt of the south, including many Negro tenant farmers. In spite of the direct channeling of this aid to the most needy, however, the financial resources made available to it have been very small relative to production control and loan and storage programs.

Efforts to promote more efficient use of resources

Efforts to promote more efficient use of resources in agriculture are by no means new. Such efforts fall broadly into three groups.

1. For years the Department of Agriculture has conducted a broad program of scientific research and guidance for farmers, both directly in its own activities and in collaboration with other agencies. Research in plant and animal biology, in farming methods and farm management, in home economics, in engineering, in weather conditions and forecasting, in economics and sociology, and in many other fields has been steadily carried on both within the Department and through a wide system of Agricultural Experiment Stations connected with the land-grant colleges. The results of this research, partially or completely financed by federal funds, are made available to farm families through publications and direct contact, especially through the nationwide system of county agents, as well as indirectly via other channels. Probably no other economic group has received such widespread assistance of this sort as has agriculture.

This research program has in some respects fitted in very effectively with other aspects of the agricultural program, but it has sometimes also been in direct conflict with that program. For example, when AAA sought to restrict corn production by restricting acreage, and at the same time the agricultural experiment stations produced hybrid corn yielding many more bushels per acre, the conflict is apparent.

2. A central part of the efforts to promote more efficient use of agricultural resources has been the soil conservation program. Programs for soil conservation are often primarily educational, giving the farmer research findings on how to operate his farm most efficiently for his own long-run interests as well as for those of all society. But there has been considerable confusion in recent years in the apparent alliance between the program of soil conservation and production control. In some instances the short-run aim of restricting agricultural output is consistent with the long-run aim of conserving natural resources, where such restriction involves shifts from soil-depleting to soil-building crops. But a large part of soil conservation is most effectively accomplished in other ways, and the long-run results of such a program are of course to increase the productive capacity of the nation's farm plant, or at least to retard its decrease in productivity.

That a great amount of unnecessary wastage of natural resources continually occurs, no one would deny. Partly this stems from inadequate information as to the best ways to use our soil; partly from the desire of many entrepreneurs to prosper now at whatever future cost; partly from the failure of the government to visualize long-period land-use needs; partly from the poverty of a vast number of farmers, which forces them to supplement skimpy incomes by "using up" the soil without replacing it by proper conservation methods. That some action to protect society in these respects is desirable is probably now generally accepted. Indeed, in the last decade the United States has become "conservation conscious"—soil conservation has become a much-hallowed phrase and great emphasis is placed on saving natural resources for future generations. Unfortunately, this attitude can easily be, and has been, much overdone. Any amount of money spent on conservation is now likely to be regarded as a good thing. But it would obviously be just as foolish to save all our resources for our grandchildren as to exhaust them all now. How the decision between present and future should be made when government steps in to undertake conservation activities the economist cannot say. But he can emphasize the need for careful marginal weighing of the alternatives, economic and social, rather than blind assumption that all conservation activities are "good" or "bad" per se.

It must not be thought, however, that the justification of our conservation programs rests entirely on provision for future years. On the contrary, equally if not more important are the beneficial results for the present of an intelligent conservation program. Among these are: (a) Provision of information to farmers concerning methods of cultivation that both conserve the soil and improve the farmers' position. Frequently

farmers fail to practice effective soil conservation even when it would be immediately economical for them to do so. (b) Regulation of farming practices in situations where the social costs of particular types of action are greater than private costs. It may be advantageous to the individual farmer to cut away a wooded space, but this may involve large additional social costs if it intensifies the problems of flood control or destroys the natural beauty of the territory. Under such circumstances free market forces fail to accomplish results consistent with the general welfare. (c) Collective action of various sorts in order to attain particular social objectives that cannot be accomplished by individual action. Preservation of city areas for parks and playgrounds is an example. Here the problem is the group one of deciding which are the most important ends sought by society and how much sacrifice is justified to attain any one of them.

3. An entirely different set of problems concerns the efficient use of resources over long periods where from year to year there are unpredictable and erratic changes in production conditions, crop yields, demand conditions, farm prices, and farm incomes. Inaccurate anticipations on the part of entrepreneurs are a major cause of inefficiency and waste in the use of agricultural resources.

As one step to minimize this inefficiency, in 1938 "all-risk" crop insurance for wheat farmers became a reality after long years of discussion and experimentation. This step to remove the uncertainties of natural forces is still in a highly experimental stage. At present the Federal Crop Insurance Corporation receives premiums in wheat, holding the premiums as a reserve from which to make payments in case production fails. Insurance is on production only, and may be taken for either 50 per cent or 75 per cent of average yield. The crop insurance program is an obvious corollary of the ever-normal granary program and of the general program to facilitate accurate planning, but how far it will be extended to other crops will depend upon experience on the now limited scale of operations.

Such a program is easily justified if its costs are less than the wastes that it eliminates. Moreover, if programs of this sort prove workable, society gains through increased stability of available farm products. The marginal gain to consumers in having more in years of shortage is probably greater than their sacrifice resulting from the withholding of crops in years of surplus. Farmers, meanwhile, gain through greater stability of income.

Conclusion

The complex of economic and broader social problems present in agriculture, together with the variety of measures already adopted for alleviation of the difficulties faced, should make it apparent that evaluation of any particular piece of farm legislation is not an easy task, much less that of evolving from the complexity of the problem an overall set of "satisfactory" corrective measures that would have any chance of both

political adoption and practical success. The application of economic tools to complex social problems is a far more difficult task than is suggested by the simple analytical problems posed in elementary texts. Especially difficult is the task of keeping separate "economic" and ethical value judgments when discussing such a broad problem, or more accurately, set of problems. Yet it is worth while in conclusion to make a tentative evaluation of the overall farm program of the last decade with suggestions for a program more "economic" and more effective in meeting the fundamental issue of income redistribution. The fact that this evaluation and the suggestions made reflect to some extent the personal inclinations of the authors is inevitable. Here as in most questions of social policy the formal tools of economic analysis are not alone sufficient, and seldom can their use be entirely colorless on such problems. It is necessary to be continually on guard against hidden personal biases even in conclusions that attempt to be entirely open and unarbitrary.

In perspective, two fundamental facts stand out about the farm legislation of the past decades. One is the success of farm groups in obtaining legislation with the stated purpose of elevating relative farm prices substantially above the level dictated by the trend of economic adjustments. The other, closely related, is the failure to separate the "economic" problem of resource allocation from the "ethical" one of income distribution.

There can be little doubt that the agricultural sector of our economy has been "economically" overexpanded and that insofar as resource allocation is concerned our endeavors should have been directed to expediting the flow of resources out of agriculture into occupations where they could support themselves without subsidies. The fundamental change in the export market and the vast relative expansion in industrial and commercial areas make this conclusion inescapable. Even under war conditions agricultural incomes have kept pace only with special aid. The concepts of "parity price" and "parity income" are based upon the assumption that all existing farmers "ought" to be again relatively as well off as they were in the period of greatest agricultural prosperity—that this income level "should" be maintained without the withdrawal of farmers to other occupations. This assumption simply does not withstand analytical examination. Programs such as the AAA, which artificially raise farm prices and *make it necessary* for resources to remain in agriculture if they are to receive the government supplementary income that they sorely need, not only do not facilitate movement of resources in accordance with consumer demand but on the contrary impose direct barriers to such a movement. Moreover, the present acreage control system brings the withdrawal of a certain per cent of all farm land, good and bad, from production, as each farmer reduces to meet his acreage allotment. From the point of view of efficiency in production the withdrawal of part of our best land from production while simultaneously subsidizing marginal land to keep it in production is clearly absurd. Economic considerations surely indicate the desirability of maintaining full production on the better land while withdrawing marginal land completely. The similarity

to the "excess capacity" wasted under the cartel arrangement in industry in spite of monopolistically high prices is striking.

If it is argued that everyone ought to receive a certain minimum income (an ethical judgment), since many farmers receive very low incomes it follows that steps "should" then be taken to augment low farm incomes. But this problem of income distribution would most logically be treated directly on an *ad personam* basis, rather than by forcibly perpetuating an overexpanded agricultural "industry" relative to other sectors by making aid to farmers contingent upon their staying in farming but reducing individual outputs. Consideration of the results of the farm program in this connection is instructive. Aid to farmers has been based not on personal need but on the size of the agricultural plant previously operated by the farmer. Thus the larger the farm, the larger have been the payments for holding a certain percentage of the acres out of production. What evidence there is available indicates that the average amount of government payments received per farm family *increases* substantially as family income *rises*. That is, as a result of the particular methods now being used for distributing payments, the largest government subsidies are going to those farm families already receiving the largest incomes.

If one accepts the dual propositions that the most "economic" allocation of resources possible ought to be obtained and that special steps ought to be taken to augment the income of low income farm families (as part of a general program for all low income groups), a logical separation of steps appears to follow. On the one hand, barriers ought to be removed and positive steps taken to expedite the free movement of resources to those areas where consumers want them most. On the other, given this economic allocation of resources, direct independent action ought to be taken to redistribute income so as to augment the means of the low income groups.

More efficient use of resources within agriculture and out of it might be obtained by removing impediments to mobility resulting from the present system of farm aid, by providing better information as to available employment opportunities, by special grants and free training to persons wishing to move from declining to growing occupations, by breaking down barriers now existing in industry as a result of industrial and labor monopoly practices, by taking measures to minimize uncertainty in agriculture, by taking steps to minimize depressions that wipe out the demand for resources everywhere, and by continued and expanding programs of soil erosion control, forest preservation, flood control, and so on.

The problem of income distribution basically is a personal one. Given the optimum economic allocation of resources, income may be redistributed by any of a number of means. Most obvious and direct is the imposition of progressive taxes on high income groups with the proceeds used to provide free or subsidized services or even cash grants to low income groups. The possibilities of such services for low income farm families fairly shout—education, better housing, rural electrification, and so on. FSA has already given some indication as to what can be

accomplished along this line. Free vocational education may be a means of solving some low income problems as well as of directing resources into the channels where they will be most productive. And positive, vigorous efforts to raise low farm incomes would go far toward solving the conservation problem. In a vast number of cases, the basic reason for soil-wasting practices is the inability of the farmer to afford soil-conserving methods that would pay both him and society in the long run.

Granting the logical validity of this general approach to solution of "the farm problem," one might proceed to detail the problems involved in its application. Can industrial and labor monopolies be broken down to permit free entrance of surplus farm population? Is it sociologically desirable that persons and families break firmly established ties to move about, even granting the economic desirability of such mobility? May it be better from a sociological and psychological viewpoint in depression periods to provide relief for farm families by making payments in relation to crops rather than simply handing out relief funds? How can we prevent depressions so as to keep open opportunities for employment of resources outside agriculture? How can the public and Congress be convinced that such fundamental changes ought to be made, here or in the field of labor-security legislation where the resource allocation and income problems have been similarly confused? How can "objective" analysis of needs and likely results be made the major guide to legislative and administrative action, instead of the present political catering to self-seeking, if perhaps well meaning, political pressure groups? The methods of formal economic analysis have much to contribute to a more reasoned formulation of social policies. But the problems of the real world cannot be solved by formal analytical tools alone—it would be an incomparably simpler world if they could.

INDEX

Index

A

"Ability-to-pay" principle of tax allocation, 736 ff.

Adjustments:

agricultural production, 888 ff.

international:

under fluctuating exchanges, 853 ff.

under international gold standard, 851 ff.

labor disputes, 496 f.

long-run, 263

joint products, 259 f.

monopolistic competition, 297

monopsonistic competition, 300

oligopoly and oligopsony, 309 f.

profit motive and elimination of pure profits in, 547 f.

marginal:

analysis of, in terms of outputs, 193 ff.

introduction to, 170 ff.

maximum profit, crudeness of approximations to, 307 f.

marginal input, graphic representation of, 190 ff.

output, long-run price and, 225 ff.

production, marginal analysis of, 181 ff.

short-run market, and the firm, 216 ff.

short-run price, monopolistic competition, 296

toward competitive equilibrium, 208 f.

Administered prices, defined, 90

Advances, federal reserve system, 647

Advertising:

basis of monopolistic power, 333 f.

effects of, on consumer selections, 332 f.

estimating effects of, 317 f.

increasing demand by, 316

price-cutting and quality competition versus, 320 f.

Age differences in income levels, 531 f.

Agents of production, defined, 78

Agreements, collective:

as legal contracts, 491

contents of, 458 ff

Aggregate demand, 45 f.

curves, downward slope of, 51 f.

Agricultural Adjustment Act:

marketing quotas under, 903 ff.

pricing under, 391

production control under, 903 ff.

Agricultural enterprise, trends in size of, 85

Agriculture:

a case study, 883 ff.

business fluctuations and, 680 f., 893 ff.

changing fortunes in, 885 ff.

claim of, to special aid, 897 ff.

farm aid programs since 1933, 901 ff.

farmer's economic status in society, 883 ff.

international trade and, 895 ff.

lack of concentration in, 94 f.

production, adjustment of, 888 ff.

Allocation, see Resources, allocation of
Allocations, in wartime, 801 ff.

American Federation of Labor, formation of, 449

Analysis, economic, see Economic analysis

Antitrust Division, Department of Justice, positions taken by, 500 ff.

Antitrust labor cases, recent court decisions in, 498 f.

Antitrust laws:

business cycle policy and, 698 ff.

enforcement of, 404 ff.

federal, 401 ff.

preservation of competition, 407 ff.

Area, income distribution by, 530 f.

Assessments, special, 786 ff.

Auxiliary functions, of middleman, 33 f.

Average expenditure:

defined, 152

graphic representation of relation between marginal cost and, 174

measure of "worth" of unit of input in other uses, 174 f.

purely competitive and monopsonistic buyers, 172 ff.

- Average product:**
 defined, 111
 graphic presentation of, 115
Average revenue:
 defined, 152
 relation between margin unit revenue
 and, 194 f.

B

- Balance of payments, international trade,**
 839 ff.
Balancing of processes, affecting size of
 enterprise, 86
Bank credit, *see* Credit
Banks:
 commercial:
 contraction of credit by, 597 ff.
 creation of credit by, 589 ff.
 Federal Reserve, 636 ff.
 individual, limits to creation of credit
 by, 593 ff.
 middlemen between savers and bor-
 rowers, 586 ff.
 quantity of credit and, 586 ff.
 see Business cycles
Bargaining, collective:
 attitudes toward, 456 ff.
 employers' policies and weapons in,
 463 ff.
 guarantee of right to, 491 f.
 right to, 490 ff.
 union policies and weapons in, 461 f.
 wage and hour policies in, 471 ff.
Basing-point pricing, 388 ff.
"Benefit" principle, of tax allocation,
 735 f.
Benefits of unemployment insurance,
 817 f.
Blacklist, employers', 464
Blocked exchange, 850 f.
Bondholders, functions of, 68
Borrowing:
 as source of national revenues, 792 ff.
 long-run effects of, 797 ff.
Bottlenecks:
 controlling and creating, 366 f.
 marketing, regulation of, 430 f.
 problem of, restraint of trade, 411 f.
Boycott, right to, 494
Business:
 freedom in choice of, "end" of social
 policy, 142
 specialization in, advantages of, 23
Business activity:
 general, widespread fluctuations in, 548
 level of. entrepreneurs and, 660 ff.

- Business cycles:**
 depression, 664 ff., 679
 economic policy to mitigate, 682 ff.
 fiscal policy to mitigate, 790 ff.
 framework of, 661 ff.
 meaning of, 657 ff.
 policy, in wartime, 703 f.
 prosperity, 670 ff., 679
 recession, 675 ff., 679
 revival, 667 ff., 679
Business enterprise:
 factors influencing size of, 85 ff.
 government-operated, concentration of,
 91 ff.
 management and size of, 87 ff.
 marketing opportunities and size of,
 89 f.
 purely competitive, maximum profit
 level of production in, 152 ff.
 size of, in U. S., 79 ff.
 technology and size of, 85 f.
 unit of, 61 ff.
Buyers:
 monopsonistic:
 facing monopolistic seller, 167 f.
 marginal costs and average expendi-
 tures of, 173 f.
 selling in purely competitive market,
 165 ff.
 selling monopolistically, 167
 production adjustments of, 188 ff.
 selling purely competitively, produc-
 tion adjustments of, 186 ff.
 purely competitive:
 marginal costs and average expendi-
 tures of, 172 f.
 selling monopolistically, production
 adjustments of, 184 ff.
Buying:
 monopsonistic competition in, 135 f.
 monopsony in, failure of marginal out-
 put analysis to discriminate be-
 tween pure competitive and,
 195 f.
 oligopsony in, 136 f.
 pure competition in, 133 f.
 failure of marginal output analysis
 to discriminate between monop-
 sony and, 195 f.
 simple monopsony in, 134 f.
By-products, utilization of, 101 f.
 affecting size of enterprise, 86 f.

C

- Capital gains and losses, 558 f.**
Capitalization, 783

Cartelization:

- governmentally sponsored, recent price policy under, 390 ff.
- publicly fostered, 364 ff.
- recent, under federal auspices, 365 f.

Cartels, trade associations as basis of, 362 f.

Chain store, trends in size of, 83 f.

Checks, clearing and collection of, 640 f.
see Credit

Children, dependent, aid to, 821

Choice, freedom in, of job or business, as "end" of social policy, 142, 714

Circular flow of economic activity, 37 f.

Citizens, sovereignty of, 716 ff.

Clayton Act, restraints of trade, 403

Coal, bituminous, pricing of, 393 f.

Codes, "fair competition," and NIRA, 364 f.

Collective agreements:

- as legal contracts, 491
- contents of, 458 ff.

Collective bargaining:

- attitudes toward, 456 ff.
- employers' policies and weapons in, 463 ff.
- guarantee of right to, 491 f.
- right to, 490 ff.
- union policies and weapons in, 461 f.

Commodity Credit Corporation, crop loans and storage under, 905 f.

Common law, restraints of trade and, 398 ff.

Common stockholders, 66 f.

Communities, economic cooperation of, 20 f.

Comparative advantage, law of, 23 f.
international trade, 830 ff.

Comparative costs, law of, free trade, and the price system, 835 f.

Compensation, workmen's, 821 ff.

Competition:

- character of, and product diversification, 334 ff.
- demand creation, 318 ff.
- indirect price, 308 f.
- meaning of, in application of antitrust laws, 408 f.
- monopolistic, 294 ff.
 - effect of excise taxes on, 748 f.
 - in selling, 135 f.
- monopsonistic, 298 ff.
 - in buying, 135 f.
- non-price, among sellers, 313 ff.
- price, nonaggressive, among oligopolists, 303 ff., 306 f.
- pure:
 - allocation of resources under, 277 ff.

Competition (Cont.):

pure (Cont.):

- effect of excise taxes on, 744 ff.
- in buying and selling, 133 f.
- law of supply and demand and, 207 ff.
- position of joint product firm and, 254 f.
- production policies of firms, 152 ff.
- quality, 296 f., 314 ff.
- effects of, 329 ff.

Competitive equilibrium, adjustment toward, 208 f.

Concentration, industrial, relation between size and, 90 ff.

Conditions of sale, unreasonable requirements concerning, 377 f.

Conflict of interests, international trade and, 837 f.

Congress of Industrial Organizations, formation of, 453

Consolidations:

- business, 358 f.
- industrial, size of, 82 f.

Construction, concentration in, 95

Consumer demand, 43 ff.

- errors in judgments of, 332

Consumers:

- economic system and, 16 f.
- preferences of, 271 ff.
 - allocation of resources in accordance with, 141 f.
 - how free price system registers, 36
- sale value of marginal product, as measure of "worth" to, 180
- sovereignty of, 716 ff.

Consumer services, concentration in, 95 f.

Consumer spending, effects on distribution of income, 536

Consumption:

- excise taxes, effects on, 749 ff.
- family expenditures for, tables of, 8-9
- production for, 4 ff. •
- retail sales taxes, effect on, 755

Contracts:

- freedom of, 489
- in restraint of trade, 399
- legal, collective agreements as, 491

Control:

- exchange rates, 877 f.
- production, under AAA, 903 ff.

Controlled exchange rates, 849 ff.

Cooperative, the, as business unit, 75

Cooperative selling agencies, 363 f.

Corporation income taxes, 771 ff.

- evaluation, 775 f.
- incidence, 773 ff.

Corporations:

- advantages of, 70 f.

Corporations (Cont.):

- as business unit, legal form of, 65 f.
- control over large investments, 72 ff.
- disadvantages of, 71
- financing of enterprise, 66 ff.
- in organization of close-knit federation, 351 f.
- prevalence of, 71 f.

Costs:

- comparative, law of, 835 f.
- consumer preferences and, 274 ff.
- decreasing, relation between, and simple monopoly, 284 f.
- explicit and implicit, 117 f.
- farmer's view of, 121 f.
- fixed:

- factor in agricultural production adjustment, 889 ff.
- farm labor and, 891 ff.
- variable and, short-run distinction between, 122 ff.

fixed unit, defined, 124

- relation to total unit, 125

joint, and joint products, 253**long-run supply and, relation between, 232 ff.****long-run view of, 128 ff.****looking ahead at, 227 f.****lowest total unit, long-run uniformity of, under free competition, 232 f.****manufacturing corporation's view of, 118 ff.****marginal:**

- defined, 171

- graphic representation of relation between average expenditure and, 174

- purely competitive and monopsonistic buyers, 172 ff.

marginal unit, compared with marginal unit revenue to find most profitable output, 196 ff.

- defined, 194

- relations between, and other cost data, 200 ff.

meaning of, 117 ff.**opportunity, of marginal products, 275 ff.****output curves, presentation of, 200 ff.****readjustment, 338****relation between, and size of industry, 235 f.****selling, meaning of, 316 f.****shifting employment, 341 f.****short-run view of, 122 ff.****"sticky," in business cycles, 663 ff.****Costs (Cont.):****total:**

- entry or withdrawal of firms and, 131 f.

- position of firm, in short run and, 157 ff.

total unit:

- "decreasing costs" and, 284

- defined, 124 f.

- least-cost-combination and, 128 f.

- long-run supply and, 233 ff.

- relation between marginal unit costs and, 200

- relation to fixed unit and variable unit, 125

variable:

- defined, 122

- position of firm, in short run and, 159

variable unit:

- defined, 124 f.

- relation between marginal unit costs and, 200

Court decisions, recent, in antitrust labor cases, 498 f.**in public utility cases, 420 ff.****Coverage of unemployment insurance, 816 f.****Credit:****bank:**

- federal reserve control of, 687 f.

- Treasury control of, 686 f.

circulation of, velocity, 621 ff.**contraction of, by commercial banks, 597 ff.****control by federal reserve, 643 ff.****creation of, by banking system, 596 f., 589 ff.****expansion and contraction of in business cycles, 668 ff., 676 ff.****installment, control of, by federal reserve system, 650 f.****introduction to, 579 ff.****limits to creation of, by individual banks, 593 ff.****money and, 581 f.****quantity of, 619 ff.****types of, 584 f.****wartime policy, federal reserve system, 651 f.****Creditors, position of, price level changes and, 607 ff.****Crop loans and storage, under CCC, 905 f.****Cycles, see Business cycles**

D

- Death taxes, 768 ff.
 - evaluation, 770 f.
 - incidence, 770
- Debs, Eugene, 464
- Debt, government, fiscal policy and, 697 f.
- Debtors, position of, price level changes and, 609 ff.
- Demand:
 - aggregate, 45 f.
 - curves, downward slope of, 51 f.
 - changes in, 47 ff.
 - effect on equilibrium price and quantity, 210
 - one product relative to others, 519
 - consumer:
 - defined, 43
 - errors in judgment of, 332
 - producer response to, 60
 - curves, downward slope of, 50 ff.
 - derived, for productive service, 239 ff.
 - elasticity of:
 - defined, 52 ff.
 - factors affecting, 57 ff.
 - graphical analysis, 55 f.
 - importance of, in setting of utility rates, 423
 - significance, 56 f.
 - farm, inelastic, production curtailment and, 893
 - graphic representation of, 45 ff.
 - increase in, response to, 228 ff.
 - individual, 44 f.
 - curves, downward slopes of, 50 f.
 - interdependent, analysis of, for productive services, 248 ff.
 - interrelated:
 - for different things, 59 f.
 - for services of different productive agents, 243 ff.
 - long-run, for productive service, 239 ff.
 - purely competitive, for productive services, 220 ff.
 - short-run, aggregate, 223
 - supply, *see* Supply and demand
- Demand creation, 316 ff.
 - effects of, 332 ff.
- Demand schedules:
 - defined, 207
 - for joint products, 257 f.
 - individual firm, 220 ff.
- Democracy, American, evolving concepts of individualism in, 346 f.
- Dependents, aid to, 821
- Depressions:
 - business cycle, 664 ff., 679

Depressions (*Cont.*):

- economic policy to mitigate, 682 ff.
- expenditure policies, government finance, 694 ff.
- revenue sources, government finance, 692 ff.
- Differentiation, product, and product diversification, 336
- Directorates, interlocking, 364
- Discrimination:
 - against particular groups of people, 342
 - price, 169
- Diseconomies of scale of enterprise, 129
- Disputes, labor, 461 ff.
 - adjustment of, 496 f.
- jurisdictional, between rival unions, 469 f.
- Distribution:
 - family income, in U. S., 10
 - government expenditures, 725 ff.
 - income, in U. S., 525 ff.
 - economic contribution and, 561 ff.
 - equity in, 483 ff., 714
 - as "end" of social policy, 142 f., 714
 - monopolistic and monopsonistic elements in, 552 ff.
 - public policy and, 561 ff.
 - relation between, and resource allocation, 533 ff.
 - tax burden, 738 ff.
 - under free price system, 37
- Diversification:
 - advantages of, 24 f.
 - products:
 - causes of, 98 ff.
 - character of competition and, 334 ff.
 - relation between leadership of firm and, 96 ff.
- Division of labor, *see* Labor, division of
- Dollar, hypothetical, path of, through the firm, 537 ff.
- Dollar votes, income distribution as determinant of, 272, 533
- Duplication, wasteful, in quality competition, 330 f.

E

- Economic activity, circular flow of, 37 f.
- Economic analysis:
 - in analysis of social organization, 17 f.
 - institutional framework of, 147 ff.
 - see also* Economic theory
- Economic contribution:
 - allocation of resources and, 564 f.
 - distribution of income and, 561 ff.
 - economic opportunity and, 564

Economic contribution (Cont.):

equity in income distribution and, 142 f.
payments to productive agents and, 327 f.

purely competitive general equilibrium and, 562 f.

uncertainty and, 563 f.

Economic opportunity:

economic contribution and, 564

entrepreneurial mobility and, 337 f.

labor mobility and, 339 f.

Economics, role of, 144 ff.**Economic system:**

defined, 13 f.

individuals and, 16 f.

place of, in achievement of standard of living, 14 ff.

Economic theory:

and social policy, 140 ff.

method of, 41 f.

Economies of scale of enterprise, 129**Economist, the:**

clarification of social ends and, 145 f.

facts, selection and interpretation of, and, 146 f.

means, selection of, by, 147

see Economic analysis, Economic theory

Economizing, problem of, 3 ff.**Elasticity of demand:**

factors affecting, 57 ff.

graphical analysis and, 55 f.

meaning of, 52 f.

significance of, 56 f.

Employees, number in manufacturing establishments, 80 f.**Employers:**

attitudes and arguments of, collective bargaining, 457 f.

law and weapons of, 494 f.

legal powers of, 488 ff.

policies and weapons in collective bargaining, 463 ff.

Employment:

cost of shifting, 341 f.

excise taxes, effect on level of, 751 f.

expenditures in periods of full, 803 f.

full-time, equivalent, table, 28

payroll taxes, effects on, 756 ff.

personal income taxes, effects on, 767 f.

provisions concerning, 458 f.

retail sales taxes, effect on, 755 f.

taxation in periods of full, 791 f.

Enforcement:

antitrust laws, and uncertainty, relation between, 406 f.

provisions for, collective agreements, 461

Enterprise:

business, see Business enterprise

scale of, of simple monopoly, 288 f.

Entrepreneur:

government as, 724 f.

level of business activity and, 660 ff.

Entrepreneurial mobility, economic opportunity and, 337 f.

Entrepreneurial preferences, 338

Entrepreneurship, defined, 77 f.

Entry:

freedom of, long-run adjustments with, 309 f.

of firms, total cost and, 131 f.

to industry, restrictions on, 283, 310, 339

to labor union, restrictions on, 467 ff.

Equality, need and, 565 ff.

Equalization of income distribution, arguments for and against, 566 f.

"Equalize costs of production" argument for protective tariff, 864 ff.

Equation of exchange, 612 ff.

Equilibrium:

competitive, adjustment toward, 208 f.
general:

characteristics of, 267 f.

usefulness of concept, 268 ff.

long-run, adjustment toward, 226 ff.

purely competitive general:

economic contributions and, 562 f.

property income and, 542 ff.

Equipment, expensive, economy in use of, 86

Escape devices, in lieu of open price competition, 321

Establishment:

defined, 78

sizes of, 79

Evasions, price-control, among sellers, 313 ff.

Exchange:

advantages of, 22 ff.

as marketing function, 31 f.

clearing, controlled exchange rates in, 851

control, 877 f.

economy:

functions of middlemen in, 30 ff.

modern, 19 ff.

equation of, 612 ff.

rates:

defined, 846

controlled, 849 ff.

fluctuating, international adjustments under, 853 ff.

"free" currencies, 847 f.

gold standard, 848 f.

- Exchange (*Cont.*):
 rationing, 850 f.
 relation between specialization and, 21 f.
- Excise taxes:
 incidence of, 743 ff.
 evaluation of, 749 ff.
- Expectations, importance of in business cycles, 661
- Expenditure:
 average:
 defined, 152
 graphic representation of relation between marginal cost and, 174
 measure of "worth" of unit of input in other uses, 174 f.
 of monopsonistic buyers, 173 f.
 of purely competitive buyers, 172 f.
 family, for consumption, tables of, 8-9
 federal, state, and local, 708 ff.
 government, 803 ff.
 distribution of, 725 ff.
 by income groups, 731 f.
 historical trend of, 711 ff.
 income redistribution and, 571 f.
 in recent years, 708 ff.
 regular, 725 ff.
 scope of, 720 ff.
 war, 727 ff.
 pattern, by groups, 533 ff.
- Expenses, operating, estimating, 423
- Explicit costs, 117
- Export subsidies, 876 f.
 in agricultural program, 907 f.

F

- Facts, selection and interpretation of, by economist, 146 f.
- Fair Labor Standard Act, *see* Wages and Hours Law
- Family:
 economic cooperation of, 20
 income distribution by, 527 ff.
 large farm, 898
- Farm aid programs:
 direct aid to farmers, 908 f.
 1921-1933, 900 f.
 since 1933, 901 ff.
- Farmers:
 direct aid to, 908 f.
 economic status of, in society, 883 ff.
 economic system and, 16
 income status of, 897 f.
- Farm labor, associated fixed costs and, 891 ff.
- "Favorable balance of trade" argument, for protective tariff, 861 f.
- Federal Deposit Insurance Corporation, function of, 689 f.
- Federal reserve system:
 control of bank credit, 687 f.
 credit control by, 643 ff.
 organization of, 636 ff.
 rediscount rate changes, 645 ff.
 service functions of, 638 ff.
- Federal-state-local fiscal activities, integration of, 806 ff.
- Federal Surplus Commodities Corporation, surplus disposal program of, 906 f.
- Federal Trade Commission Act, and restraints of trade, 403
- Federations:
 close-knit, 351 ff.
 loose-knit, privately organized, 359 ff.
- Finance, government, 690 ff.
- Financial manipulation, affecting size of enterprise, 90
- Financiers, outside, speculative manipulation of prices by, 383 f.
- Financing:
 affecting size of enterprise, 90
 as function of middleman, 33
 corporate enterprise, 66 ff.
 international transactions, 845 ff.
 unemployment insurance, 817
- Firms:
 defined, 77
 entry or withdrawal of, total costs and, 131 f.
 exclusion of new, as foundation for simple monopoly, 283
 freedom and restrictions of entry of, in long run, 137 ff.
 in market, 133 ff.
 input-output relations in, 106 ff.
 joint products:
 output policy of, 255 f.
 position of, pure competition and, 254 f.
 leadership of, relation between diversification of products and, 96 ff.
 monopolistic, production and price adjustments of, 160 ff.
 monopsonistic, short-run policies of, 165 ff.
 number of, in an industry, 90 ff.
 position of:
 total unit costs, in short run, 157 ff.
 variable unit costs and, in short run, 159
 purely competitive:
 evaluation of level of production in, 182 ff.

Firms (Cpnt.):

purely competitive (*Cont.*):

graphic representation of profits and losses of, 159 f.

marginal analysis of production adjustments, 181 ff.

production policies of, 152 ff.

short-run market adjustments and, 216 ff.

supply schedule of an individual, 216 f.

Fiscal activities, federal-state-local, integration of, 806 ff.

Fiscal policy, 690 ff.

government debt and, 697 f.

national income and, 790 ff.

timing of, 696 f.

Fixed costs:

defined, 122

factor in agricultural production adjustment, 889 ff.

farm labor and, 891 ff.

per unit of output, 124

Fixed productive agents, 107

Fluctuations:

basis for economic policies to mitigate, 684 f.

business, agriculture and, 893 ff.

economic policy and, 682 ff.

exchanges, international adjustments under, 853 ff.

general business, 657 ff.

widespread, in business activity, 548

see Business cycles

Foreign exchange, 845 ff.

"Free" currencies, exchange rates with, 847 f.

Freedom of contract, 489

Free price system:

circular flow of economic activity in, 37 f.

consumers' preferences registered by, 36

distribution under, 37

production under, 36 f.

simplified model of, 35 ff.

Free trade, law of comparative costs and, 835 f.

G

Gift taxes, 768 ff.

evaluation, 770 f.

incidence, 770

Gold:

future policy regarding, 634 f.

price level and, 627 ff.

silver, and "mixed standard," 633 f.

Gold (Cont.):

"too little," problem of, 631 f.

"too much," problem of, 632 f.

Gold standard, the, 627 ff.

abandonment of, 629 ff.

exchange rates under, 848 f.

international, international adjustments under, 851 ff.

Gompers, Samuel, 449

Government:

as entrepreneur, 724 f.

as transfer agent, 724 f.

debt, fiscal policy and, 697 f.

expenditures:

distribution of, 725 ff.

by income groups, 731 f.

historical trend of, 711 ff.

in recent years, 708 ff.

regular, 725 ff.

scope of, 720 ff.

war, 727 ff.

federal reserve system as fiscal agent for federal, 642 f.

finance, 690 ff.

in American economy, 40 f.

ownership, 427 ff.

regulation, public utility, 427 ff.

revenues:

historical trend of, 711 ff.

in recent years, 708 ff.

sponsored cartelization, recent price policy under, 390 ff.

support of collaborative pricing, 395 f.

trade-unions and, 488 ff.

Government-operated enterprise, concentration of, 91 ff.

Grademark, sellers required to use privately owned, 378

Grading of products, as marketing function of middleman, 32 f.

Great Britain, development of trade-unionism in, 444 ff.

Gross income, defined, 121

Group action, displacement of individual by, in American economy, 39 f.

H

High-fixed-cost industries, incentives to collaborative price stabilization, 384 ff.

High-price policy:

of nineteenth century trusts, 380 ff.

recent, example of, 382 f.

Holding companies:

development of, 353 ff.

utility, see Utility holding companies

Horizontal integration, defined, 82

Hours:

- legislation controlling, history of, 516 f.
- policies, economic effect of, in trade-union bargaining, 471 ff.
- provisions, in labor agreements, 460 f.

I

Implicit costs, 118

Implicit interest, 120

Improvements, real property, taxes on, 780 f.

Incidence of taxation, 741 ff., 743 ff.

Income:

- agricultural, erratic, 898 f.
- area, distribution by, 530 f.
- average per capita, 525 f.
- distribution:

- as determinant of dollar votes, 533
- by occupation, 529 f.

- consumer spending and, 536

- controlling, at source, 572 ff.

- corporation income, effect of, 775 f.

- economic contribution and, 561 ff.

- equalization of, arguments for and against, 566 f.

- equity in, 483 ff., 714, 735 ff.

- as "end" of social policy, 142 f., 714
- in socialistically organized economy, 575 f.

- in U. S., 525 ff.

- monopolistic and monopsonistic elements in, 552 ff.

- personal income taxes, effects on, 766 f.

- price level changes and, 605 ff.

- property taxes, effect on, 785 f.

- public policy and, 561 ff.

- relation between, and resource allocation, 533 ff.

- unequal, effects on allocation of resources, 535 f.

family, distribution by, 527 ff.

- distribution in U. S., table of, 10

farmers, 897 f.

groups, expenditure pattern by, 533 ff.

inequality of, measurement of, 568 f.

levels, age, sex, and race differences in, 531 f.

money, 521

- sources of, 537 ff.

- sources of personal, in U. S., 555 ff.

national, 522 ff.

- corporation income taxes, effects on, 776

- fiscal policy and, 790 ff.

Income (Cont.):

- national money, as index of national real, 524 f.

- personal, distribution of, 521 ff.

- problem of, restraint of trade, 412 f.

- property, purely competitive general equilibrium and, 542 ff.

- psychic, 522

- real, 521 f., 523

- redistribution, 533, 735 ff.

- collection of public revenues and, 570 f.

- public expenditures and, 571 f.

- public policy and, 567 ff., 735 ff.

- retail sales taxes, effect on, 755

- sources, by size of income, 559 f.

- taxes, see Corporation income taxes, Personal income taxes

- types of, 521 f.

Income groups, distribution of expenditures by, 731 f.

"Increased employment" argument for protective tariff, 868 f.

Individual:

- displacement of, by group action, 39 f.
- economic system and, 16 f.

Individual demand, 44 f.

- curves, downward slopes of, 50 f.

Individualism:

- economy and, 325 ff.

- philosophy of, 344 ff.

Industrial Workers of the World, formation of, 450

Industry:

- defined, 78 f.

- long-run under- and overdevelopment of, 327

- number of firms in, 90 ff.

- purely competitive adjustments of, 206 ff., 216 ff., 225 ff.

Inelasticity of demand, factors affecting, 57 f.

"Infant industry" argument for protective tariff, 859 f.

Inflation:

- "direct action" to check, 801 ff.

- effects on income distribution, 605 ff.

- fiscal powers to check, 690 f.

- quantity of money and, 617 f.

- wartime credit policy, federal reserve system, 651 f.

Information, as marketing function, 32 f.

Injunction:

- defined, 464

- weapon of employers, 495

Innovations, introduction of, 550 f.

- in business cycles, 668 f., 674

Input-output data:

for hypothetical firm, 112 f.

graphic presentation of, 113 f.

Input-output relations, physical, in a firm, 108 f.

Inputs, defined, 106

Installment credit, control of, by federal reserve system, 650 f.

Insurance:

old age, 818 f.

unemployment, *see* Unemployment insuranceIntegration, product diversification, as *con* lary of, 192 f.

Interest:

implicit, 129

in purely competitive general equilibrium, 544 f.

rates of, 552 f.

allocation of resources and, 546 f.

Interlocking directorates, 364

International adjustments:

under fluctuating exchanges, 853 f.

under international gold standard, 854 f.

International lending, basic case for, 836 f.

International trade:

agriculture and, 895 f.

balance of payments, 839 f.

basic case for, 829 f.

comparative advantage, law of, 830 f.

conflict of interests and, 837 f.

interferences with, 850 f., 874 f.

interregional and, 829 f.

political aspects of, 838 f.

International transactions, financing, 845 f.

Interregional trade, international and, 829 f.

Investments:

corporation, control over, 72 f.

"fair value" of, problem of, 421 f.

initial, size of, 338 f.

Investors, economic system and, 16

J

Job opportunities:

freedom in choice of, 142, 482 f., 714

scarce:

exclusion from, economic effects of, 467 f.

for agricultural labor, 892 f.

trade-unionism and collective management of, 440 f.

Job protection, provisions concerning, 438 f.

Joint products, 253 f.

characteristics of, 253 f.

Joint products (*Cont.*):

complementary and substitutive production relations between, 253 f.

demand schedules for, 257 f.

differential utility rates and, 424 f.

long-run adjustments, 259 f.

output and pricing of, 255 f., 260 f.

prices of, 262 f.

short-run pricing of, 258 f.

short-run supply of, 256

single product analysis and, 263 f.

K

Knights of Labor, formation of, 448 f.

L

Labor, *see also* Trade-unionism

disputes, 461 f., 496 f.

division of:

affecting size of enterprise, 85 f.

management, affecting size of enterprise, 87

farm, associated fixed costs and, 891 f.

groups, tactics of organized, 342

law and weapons of, 492 f.

legal powers of, 488 f.

mobility of, and economic opportunity, 339 f.

role of, in economic control, 456 f.

services of, pricing, 299

supply, controlling, 373 f.

wages of, 552

Laborers, economic system and, 16

Land, taxes on, 781 f.

Laws:

"antitrust," enforcement of, 404 f.

concerning standards, manipulations of, 377 f.

employers' weapons and, 494 f.

federal, 508 f.

"antitrust," 401 f.

interstate pacts, 507 f.

labor's weapons and, 492 f.

state, wage and hours regulation by, 506 f.

traditional principles in, 488 f.

Leadership, price, and stabilization of steel prices, 366 f.

Least-cost combination:

defined, 126

maximum profit output and, 204

Legal form:

corporation, as business unit, 65 f.

partnership, 63 f.

single proprietorship, 61

Lending, international, basic case for, 836 f.

Licensing requirements, unreasonable, 377

Living:

- agricultural standard of, 899 f.
- plane of, 4 ff.
- economic cooperation determines, 22
- present American, 7 ff.
- progress in raising, 714 f.
- as "end" of social policy, 143 f.
- trade-union policies, 485 f.
- standard of, 5 ff.
- economic system in achievement of, 14 ff.

Loans, crop, under CCC, 905 f.

Local-state-federal fiscal activities, integration of, 806 ff.

Long run:

- combinations, of productive agents, 115 f.
- economic stabilization, 701 ff.
- entry of firms in, freedom and restrictions of, 137 ff.
- output-cost curves, 129 f.
- planning, 106 ff.
- view of costs, 128 ff.

Losses, *see also* Capital gains and losses

graphic representation of:

monopolistic firm, 164 f.

purely competitive firm, 159 f.

Luxuries, defined, 749 f.

M

Machinery, expensive, affecting size of enterprise, 86

Management:

- efficiency, minimum wage regulations and, 512
- size of business enterprise and, 87 ff.

Manipulation, financial, affecting size of enterprise, 90

Manpower, in U. S., 13

Manufacturers, economic system and, 16

Manufacturing:

- concentration in, 93 f.
- corporation, cost and profit viewed by, 118 ff.
- establishments, trends in size of, 80 ff.

Marginal adjustment, analysis of:

- in terms of inputs and associated outputs, 181 ff.
- in terms of outputs, 193 ff.
- introduction to, 170 ff.

Marginal cost:

- defined, 171
- graphic representation of relation between average expenditure and, 174
- of monopolistic buyers, 173 f.
- of purely competitive buyers, 172 f.

Marginal product:

- defined, 111, 171
- graphic presentation of, 113 f.
- sale value of, as measure of "worth" of marginal product to consumers, 180
- marginal revenue and, 177 ff.

Marginal revenue:

- defined, 171
- elasticity of average revenue schedule and, 175 ff.
- sale value of marginal products and, 177 ff.

Marginal unit cost:

- compared with marginal unit revenue to find most profitable output, 196 ff.
- defined, 194
- relation between, and other cost data, 200 ff.

Marginal unit revenue:

- compared with marginal unit cost to find most profitable output, 196 ff.
- defined, 194
- relation between average revenue and, 194 f.

"Margin" requirements, changes in, federal reserve system, 649 f.

Market:

- adjustments:
 - short-run, and the firm, 216 ff.
 - toward competitive equilibrium, 208 f.
- firm in the, 133 ff.
- labor, *see* Labor
- monopsonistic, and payments to owners of productive agents, 329
- operations, open, by federal reserve, 643 ff.
- opportunities, size of business enterprise and, 39 f.
- positions:
 - and antitrust, 407 ff.
 - dominant and controlling:
 - attaining through product diversification, 103
 - foundations of, 350 ff.
 - types, classification of, 133 ff.
- prices, defined, 90
- pure competition in, 133 f.
- purely competitive, monopolistic buyer selling in, 165 ff.
- simple monopoly in, 134 f.
- see also* Competition, Monopolistic, Monopoly, Monopsonistic, Monopsony, Oligopoly, and Oligopsony

- Marketing:**
 bottlenecks, regulation of, 430 f.
 channels, controlling, 369 ff.
 functions of middlemen, 31 ff.
 quotas under AAA, 903 ff.
- Marxian socialism,** 441 f.
- Materials,** unreasonable requirements concerning, 377 f.
- Means,** selection of, by economist, 147
- Meat-packing industry:**
 controlling marketing channels in, 369 f.
 price discrimination and product diversification in, 375 ff.
- Mechanics' Union of Trade Associations** in Philadelphia, 448
- Middlemen:**
 functions of, in complex exchange economy, 30 ff.
 necessity for, 34 f.
 quantitative importance of, 34
- Milk, fluid,** pricing of, 392 f.
- Minimum wage,** *see* Wage, minimum
- Mining,** concentration in, 95
- "Mixed standard,"** gold, silver, and, 633 f.
- Mobility,** *see* Entry, and Economic opportunity
- Monetary policy,** 685 ff.
 problem of "timing," 689
- Money:**
 circulation, velocity of, 621 ff.
 credit and, 581 f.
 flow of, through the firm, 537 ff.
 functional division of, 539 ff.
 furnishing, for circulation, function of
 federal reserve system, 639 f.
 in complex society, 579 ff.
 introduction to, 579 ff.
 issue and withdrawal of, 685 f.
 new, as source of revenues, 800 f.
 quantity of:
 determinants of, 616 ff.
 inflation and, 618 f.
 transfers in public economy, 722 ff.
 types of, 582 ff.
 value of, defined, 600 ff.
- Money income,** 521, 523 f.
 personal, sources of, 537 ff., 555 ff.
- Monopolistic competition,** 294 ff.
 meaning of concept, 135 f.
- Monopolistic enterprise,** payments to owners of productive agents and, 328 f.
- Monopolistic firm:**
 graphic representation of profits and losses, 164 f.
- Monopolistic firm (Cont.):**
 production and price adjustments of, 160 ff.
- Monopolistic power:**
 advertising as a basis of, 333 f.
see also Market, positions, and Marketing
- Monopolistic sellers,** marginal products of, marginal revenue and sale value of, 175 ff.
- Monopoly, simple,** 283 ff.
 meaning of concept, 134 f.
- Monopsonistic buyers,** marginal costs and average expenditures of, 173 f.
- Monopsonistic competition,** 298 ff.
 in buying, 135 f.
- Monopsonistic firms,** short-run policies of, 165 ff.
- Monopsony, simple,** 290 ff.
 meaning of concept, 134 f.
- Motor fuel taxes,** 752 f.
- N
- "National economic stability" argument** for protective tariff, 869 ff.
- National income,** *see* Income, national
- National Industrial Recovery Act:**
 codes of "fair competition" and, 364 f.
 collective bargaining and, 491
 conflicts of interest revealed under, 321 ff.
 regulation of wages, 508
- National Labor Relations Act,** 491 f.
- National Labor Union,** formation of, 448
- National Resources Committee,** estimated value of production equipment, 12
- "National self-sufficiency" argument** for protective tariff, 860 f.
- National Trades' Union,** 448
- Natural resources,** in U. S., 12
- Need, equality and,** as social "ends," 565 ff.
- News, market,** function of, in marketing, 32 f.
- O
- Occupational groups,** selected, number of persons in, 26-27
- Occupational specialization,** in America today, 25 ff.
- Occupations:**
 income distribution by, 529 f.
 specialization, advantages of, 22 f.

Oil-refining industry:
 patents in, 373
 price cutting in, 374 f.
 controlling transportation facilities in, 367 f.

Old-age insurance, 818 ff.

Old-age pensions, 820 f.

Oligopoly, 303 ff.
 collaboration in price and output policies, 305 f.
 long-run adjustments, 309 f.
 meaning of, 136 f.
 price competition, nonaggressive, 303 ff.
 restraint of trade and, in antitrust program, 409 f.

Oligopsony, 303 ff.
 collaboration in price and output policies, 307
 long-run adjustments, 309 f.
 meaning of, 136 f.
 price competition, nonaggressive, 306 f.

Open market operations, federal reserve, 643 ff., 687 ff.

Open price competition, escape devices in lieu of, 321

Open shop, defined, 464

Operating expenses, estimating, 423

Opportunities, *see* Economic opportunity, Job opportunity

Optimum scale of enterprise, 128 f.
 divergences from, 130 f.

Organization, social, analysis of, economic analysis in, 17 f.

Outlet differentiation, basis of, 298

Output:
 costs per unit of, 124 ff.
 defined, 106
 finding most profitable, by comparing marginal unit revenues and marginal unit costs, 196 ff.
 finding most profitable, by comparing total revenues and total costs, 152 ff.
 marginal analysis in terms of, 193 ff.
 maximum profit, and least-cost-combination, 204
 volume of, management economy through, 87 f.
see also Competition, Joint products, Monopolistic, Monopsonistic, Oligopoly, and Oligopsony

Output-cost curves:
 long-run, 129 f.
 short-run, 126 f.

Output-cost curves (*Cont.*):
 oligopsony collaboration in price and, 307

P

Partnership:
 as business unit:
 advantages of, 64
 disadvantages of, 64 f.
 legal form of, 63 f.
 extent of enterprise, 65

Patents:
 as foundations of monopolistic power, 372 f.
 public utility approach to, 431 f.

Payments:
 balance of, in international trade, 839 ff.
 to productive agents and economic contribution, relation between, 327 ff.

Pay-roll taxes, 756 ff.
 evaluation, 758 f.

Pensions, old-age, 820 f.

Personal income, distribution of, 521 ff.

Personal income taxes, 760 ff.
 evaluation, 766 ff.
 incidence, 765 f.

Personal money incomes, *see* Incomes, money

Physical supply, functions of, as functions of middlemen, 33

Picket, right to, 494

Plane of living:
 defined, 4 ff.
 economic cooperation determines, 22
 present American, 7 ff.
 progress in raising, as "end" of social policy, 143 f., 714
 trade union policies and, 485

Planning, long- and short-run, 106 ff.

Plant, defined, 78

Plumbing industry, unreasonable licensing requirements of, 377

Pools, development of, 361 ff.

Power, foundations of, 351 ff.

Power sources, controlling, 371

Preferences:
 consumer, 271 ff.
 allocation of resources in accordance with, 277 ff., 713 f.
 revenues, costs and, 274 ff.
 citizens, 716 ff.
 entrepreneurial, 338
 scales of, 4 ff.
 worker, realization of, 482 f.

Preferred stockholders, 69 f.

Price: see also Pricing**adjustments:**

monopolistic firm, 160 ff.

short-run, monopolistic competition, 296

administered and market, defined, 90

agricultural, erratic, 898 f.

discrimination, 169, 374 ff.

indexes, and price level changes, 602 ff.

joint product, 262 f.

long-run, and output adjustments, 225 ff.

oligopoly collaboration in, and output policies, 305 ff.

oligopsony collaboration in, and output policies, 307

policies, examples of monopolistic:

extreme, 380 ff.

historical perspective on, 394 ff.

under governmentally sponsored cartelization, 390 ff.

speculative manipulation, by outside financiers, 383 f.

steel, price leadership and stabilization of, 386 ff.

variability in, change in demand or supply, 212

Price ceilings, evasions of, 323 f.

in wartime, 801 ff.

Price competition, nonaggressive:

among oligopolists, 303 ff.

among oligopsonists, 306 f.

Price-cutting:

advertising versus, 320 f.

discriminatory, 374 ff.

Price fixing:

difficulties of, 212 ff.

individual, 801 ff.

Price level:**changes:**

distribution of income and, 605 ff.

price indexes and, 602 ff.

defined, 600 ff.

gold and, 627 ff.

spending, trade, and interdependence of, 626

Price stabilization:

as basis for economic stabilization, 701 f.

as business policy, 384 ff.

extent of, 390

incentives to collaborative, in high-fixed-cost industries, 384 ff.

Price system, law of comparative costs, free trade, and, 835 f.**Pricing: see also Price**

basing-point, 388 ff.

Pricing: see also Price (Cont.):

bituminous coal, 393 f.

collaborative, government support of, 396 f.

fluid milk, 392 f.

interdependence in, of productive services, 250 f.

joint products, 255 ff.

labor, peculiarities of, 299

long-run, of monopolist, 287 f.

long-run considerations, of monopsonist, 292 f.

programs, privately controlled, 394 f.

services of material production agents, 298 f.

short-run, of joint products, 258 f.

short-run maximum profit, of monopoly, 285 f.

under Agricultural Adjustment Administration, 391, 902 ff.

Priorities, 801 ff.**Produced resources, in U. S., 12 f.****Producers, response to consumer demands, 60****Production:****adjustments:**

marginal analysis of, 181 ff.

of monopolistic firm, 160 ff.

agricultural, adjustment of, 888 ff.

complementary and substitutive, relation between joint products, 253 f.

control of, under AAA, 903 ff.

corporation income taxes, effects on, 776

costs of, 117 ff.

curtailment, and inelastic farm demand, 893

data, discontinuous, deriving figures for marginal unit cost and marginal unit revenue from, 193 ff.

excise taxes, effects on, 749 ff.

level of, evaluation in purely competitive firm, 182 ff.

maximum profit level of, in purely competitive enterprise, 152 ff.

most profitable level of, marginal analysis of, 181 ff.

policies of purely competitive firms, 152 ff.

specialization in, of consumers' good, 29 f.

under free price system, 36 f.

Productive agents (see also Productive services):

change in price of one, adjustment to, 246 f.

- Productive agents (*see also* Productive services) (*Cont.*):
 complementary and substitutive relations between, 245
 fixed, 107
 interrelated demands for different, 243 ff.
 long-run combinations of, 115 f.
 material, pricing services of, 298 f.
 payments to, relation between and economic contribution, 327 ff.
 short-run allocation of, 277 ff., 326 f.
 short-run combinations of, 109 ff.
- Productive services; *see also* Productive agents
 change in supplies of, 549 f.
 derived demand for, 239 ff.
 interdependent demands for, analysis of, 248 ff.
 long-run demands for, 239 ff.
 pricing of, and input under oligopsony, 306 f.
 pricing of, interdependence in, 250 f.
 short-run maximum-profit pricing of, by monopsonist, 290 ff.
 supply of, influences determining, 237 f.
- Products:
 differentiation, bases of, in monopolistic competition, 294 f.
 diversification of:
 and character of competition, 334 ff.
 causes of, 98 ff.
 relation between leadership of firm and, 96 ff.
 joint, *see also* Joint products
 differential rates and, 424 f.
- marginal:
 defined, 171
 opportunity costs and sale values of, 275 ff.
 sales value of:
 as index of preferences, 274 f.
 as measure of "worth" of marginal product to consumers, 180
 marginal revenue and, 177 ff.
 pricing of, purely competitive firms, 219 f.
 supplies of, influences determining, 236 ff.
- Profit expectations, response to, 226 f.
 in business cycles, 661
- Profits:
 defined in relation to costs, 117 ff.
 farmer's view of, 121
 graphic representation of:
 monopolistic firm, 164 f.
 purely competitive firm, 150 f.
- Profits (*Cont.*):
 long-run, of simple monopoly, 289 f.
 manufacturing corporation's view of, 118 ff.
 maximum profit level of production, 152 ff.
 of exclusive position, 553 ff.
 pure, 547 ff.
- Property income, purely competitive general equilibrium and, 542 ff.
- Property rights, 489 f.
- Property taxes, 776 ff.
 administration, 777 ff.
 capitalization, 782 ff.
 evaluation, 784 ff.
 improvements on real property, 780 f.
 real property, 781 f.
 severance, 782
- Proprietorship:
 advantages, 61 f.
 decreasing importance of individual, 62 f.
 disadvantages of, 62
 legal form, 61
- Prosperity, in business fluctuations, 670 ff., 679
- "Protect home industry" argument for protective tariff, 862 ff.
- "Protect our high wages and high plane of living" argument for protective tariff, 866 ff.
- Psychic income, 522
- Public economy:
 introduction to, 707 ff.
 money and real transfers in, 722 ff.
 social policy and, 713 ff.
 special characteristics of, 715 f.
- Public policy, income distribution and, 561 ff., 567 ff., 714 ff.
- Public utilities:
 approach, and antitrust policy, 429
 companies, privileges granted to, 418 f.
 concentration of, 91 ff.
 rate regulation, evaluation of, 425 f.
 regulation of, 415 ff.
 holding companies, 426 ff.
 machinery of, 417 ff.
 rates, 420 ff.
 sales, quality and quantity of, 419 f.
- Public utility approach:
 NIRA approach and, contrast between, 432
 to patents, 431 f.
- Public Utility Holding Company Act, 426
- Pure competition, *see* Competition, pure
- Pure profits, 547 ff.

Q

- Quality, measuring, problem of, 314 f.
- Quality competition, 296 f., 314 f.
- advertising versus, 320 f.
- effects of, 329 ff.
- Quantity, variability in, with change in demand or supply, 212
- Quotas:
 - international trade, 874 f.
 - marketing, under AAA, 903 ff.

R

- Race difference in income level, 531 f.
- Rate regulation, 420 ff.
- Rates:
 - demand elasticity in setting of, 423
 - differential, joint products and, 424 f.
 - exchange, 847 ff.
 - rediscount, federal reserve system, changes in, 645 ff.
 - see also* Wages
- Rationing, 801 ff.
- Raw materials, controlling, 371
- Readjustment, costs and difficulties of, 338
- Real estate, factor in agricultural production adjustment, 889 ff.
- Real income, 521 f.
- Real property:
 - improvements on, taxes on, 780 f.
 - taxation of, 781 f.
- Real transfers, in public economy, 722 ff.
- Recession, in business fluctuations, 675 ff., 679
- Reciprocal Trade Agreements Program, export subsidies and, 907 f.
- Rediscount rate changes, federal reserve system, 645 ff.
- Regional specialization:
 - advantages of, 24
 - in America today, 29
- Regular expenditures, government, 725 ff.
- Regulation:
 - legal, of wages and hours, 506 ff.
 - marketing bottlenecks, 430 f.
 - public utility, 415 ff.
- Rent:
 - in purely competitive general equilibrium, 544 f.
 - in income statistics, 555 f.
- Reserves, holding member bank, function of federal reserve system, 638 f.
- as means of credit control, 643 ff., 687 f.
- Resources:
 - allocation of:
 - economic contribution and, 564 f.

Resources (*Cont.*):

- allocation of (*Cont.*):
 - in accordance with consumer preferences, 141 f., 713 f.
 - monopolistic and monopsonistic interferences with, 325 ff.
 - rate of interest and, 546 f.
 - relation between income distribution and, 533 ff.
 - trade-unions and, 480 ff.
 - under pure competition, 277 ff.
 - unequal income distribution, effects of, on, 535 f.
 - war time controls of, 801 ff.
- economic, and technology, 11 ff.
- efficient utilization of, in product diversification, 100 ff.
- natural, in U. S., 12
- produced, in U. S., 12 f.
- Restraints of trade:
 - attacks on since 1938, 404 ff.
 - common law and, 398 ff.
 - oligopoly and, in antitrust program, 409 f.
 - public policy attacking, 397 ff.
 - trade-unionism and, 497 ff.
- Retailing establishments, trends in size of, 83 f.
- Retailing services, degrees of concentration in, 95 f.
- Retail sales taxes, 753 ff.
- evaluation of, 755 f.
- incidence, 753 f.
- Revenues:
 - average:
 - defined, 152
 - relation between marginal unit revenue and, 194 f.
 - consumer preferences and, 274 ff.
 - curves, presentation of, 200 ff.
 - government:
 - federal, state, and local, 708 ff.
 - historical trend of, 711 ff.
 - looking ahead at, 227 f.
 - marginal:
 - defined, 171
 - elasticity of average revenue schedule and, 175 ff.
 - sale value of marginal products and, 177 ff.
 - marginal unit, compared with marginal unit cost to find most profitable output, 196 ff.
 - defined, 194
 - relation between average revenue and, 194 f.

Revenues (*Cont.*):

- public:
 - collection of, and income redistribution, 570 f.
 - taxation, 733 ff.
 - sources of national, 791 ff.

Revival:

- business cycle, 667 ff., 679
- fiscal powers to stimulate, 691 ff.

Risks:

- as function of middleman, 33 f.
- uncertainty versus, 551 f.

Robinson, E. A. G., on machine economy, 86

S

Sale value, of marginal products, 275 ff.

Scale of enterprise:

- economies and diseconomies of, 129
- optimum:
 - defined, 128 f.
 - divergences from, 130 f.

Scales of preferences, defined, 4 ff.

Security:

- as "end" of social policy, 144, 715
- effect of trade-union policy on, 485
- quest for, in product diversification, 102
- social, 809 ff.

Sellers:

- monopolistic:
 - buying monopsonistically, production adjustments of, 188 ff.
 - buying purely competitively, production adjustments of, 184 ff.
 - facing monopsonistic buyer, 167 f.
 - marginal products of, marginal revenue and sale value of, 175 ff.
- non-price competition among, 313 ff.
- price-control evasions among, 313 ff.
- purely competitive:
 - buying monopsonistically, production adjustments of, 186 ff.
 - marginal products of, marginal revenue and sale value of, 175 ff.
- required to use grademark, 378

Selling:

- monopolistic competition in, 135 f.
- oligopoly in, 136 f.
- pure competition in, 133 f.
- simple monopoly in, 134 f.

Selling agencies, cooperative, 363 f.

Selling costs, defined, 316 f.

Seniority rules, in trade-unions, 468 f.

Severance taxes, 782

Sex differences in income levels, 531 f.

Sherman Act:

- penalties under, 407
- provisions and interpretations, 401 f.

Shifting of taxation, 741 ff.

Short-run planning, 106 ff.

- combinations of productive agents, 109 ff.
- decision to operate or shut down in, 127 f., 155
- output-cost curves, 126 f.
- preliminary analysis of, 151 ff.
- view of costs, 122 ff.

Silver, gold, and "mixed standard," 633 f.

Size, problem of, restraint of trade, 410 f.

Social criteria for taxes, 734 f.

Social order, new, trade-unions and, 441 f.

Social policy:

- "ends" of:
 - clarification by the economist, 145 f.
 - frequently accepted, 140 ff.
 - in public economy, 713 ff.
- Social security, background of, 809 ff.
- Social Security Act of 1935, 811 ff.
- Social welfare, principle of maximizing, 720 ff.

Society:

- atomistic view of, 344 ff.
- farmers' economic status in, 883 ff.

Sovereignty, consumers' and citizens', 716 ff.

Special assessments, 786 ff.

Specialization:

- advantages of, 22 ff.
- factors encouraging, in American economy, 25
- in production of consumers' good, 29 f.
- international and inter-regional, 829 ff.
- limitations of, 24 f.
- management, affecting size of enterprise, 87
- occupational, in America today, 25 ff.
- regional, in America today, 29
- relation between, and exchange, 21 f.

Spending, trade, and price level, 612 ff.

Stability:

- as "end" of social policy, 144, 715
- effect of trade-union policy on, 485

Stabilization:

- "long-run" economic, 701 f.
- price, *see* Price stabilization
- Stabilization funds, as controlled exchange rates, 849 f.

Standard of living:

- achievement of, economic system in, 14 ff.
- defined, 5 ff.

State-federal-local fiscal activities, integration of, 806 ff.
 Steel prices, price leadership and stabilization of, 386 ff.
 "Sticky" costs, 663 ff.
 Stockholders:
 common, 66 f.
 preferred, 69 f.
 Storage:
 as marketing function of middleman, 33
 crop, under CCC, 905 f.
 Strike:
 as a labor union weapon, 461
 right to, 492 f.
 Subsidies, export, 876 f.
 trade agreements program and, 907 f.
 Supply:
 aggregate short-run, 218 f.
 changes in, effect on equilibrium price and quality, 210 f.
 curve, long-run aggregate, 233 ff.
 joint product, 262 f.
 long-run, relation between costs and, 232 ff.
 periods of, 214 f.
 purely competitive, and the firm, 216 ff.
 short-run, of joint products, 256
 Supply and demand:
 changes in both, effect of, on equilibrium price and quality, 211 f.
 law of, pure competition and, 207 ff.
 Supply schedule:
 defined, 207
 long-run, 231
 Surplus disposal program, under FSCC, 906 f.

T

Tariff:
 background of, 856 f.
 development of, 857 f.
 effects of, 859 ff.
 making, in U. S., 871 ff.
 protective, arguments used to substantiate, 859 ff.
 Taxation:
 "ability-to-pay" principle, 736 ff.
 as source of revenue, 791 f.
 "benefit" principle, 735 f.
 "equity" in, 735 ff.
 excise taxes, 743 ff.
 "good" and "bad" taxes, 784 f.
 incidence of, 741 f.

Taxation (Cont.):
 public revenues, 733 ff.
 shifting of, 741 ff.
 social criteria for, 734 f.
 Tax burden, distribution of, 738 ff.
 Taxes:
 assessments, special, 786 ff.
 corporation income, 771 f.
 death, 768 ff.
 excise, 743 ff.
 gift, 768 ff.
 income, *see* Corporation income taxes,
 Personal income taxes
 motor fuel, 752 f.
 pay-roll, 756 ff.
 personal income, 760 ff.
 property, 776 ff.
 retail sales, 753 ff.
 severance, 782
 Technology:
 economic resources and, 11 ff.
 size of business enterprise and, 85 f.
 Temporary National Economic Committee, report of, on manufacturing operations in U. S., 80 ff.
 Timing of fiscal policy, 696 f.
 Timing of monetary policy, 689
 Total costs:
 entry or withdrawal of firms and, 131 f.
 position of firm and, in short run, 157 ff.
 Totalitarianism, 346
 Total product:
 defined, 111
 graphic presentation of, 113 f.
 Total unit cost:
 "decreasing costs" and, 284
 defined, 124 f.
 least-cost-combination and, 128 f.
 long-run supply and, 233 ff.
 relation between marginal unit costs and, 200
 relation to fixed unit and variable unit, 125
 Trade:
 free, *see* Free trade
 international, *see* International trade
 restraints of, *see* Restraints of trade
 spending, and price level, interdependence of, 626
 Trade agreements program, 872 f.
 and export subsidies, 907 f.
 Trade associations, development of, 362 f.
 Trade-unionism:
 allocation of resources and, 490 ff.
 bargaining, wage and hour policies in, 471 ff.
 current American, 453 f.

Trade-unionism (Cont.):

- displacement of union workers, resistance to, 470
- disputes between rival unions, 469 f.
- economics of, 467 ff.
- England and U. S., compared, 455
- government and, 488 ff.
- history and philosophy of, 439 ff., 443 ff.
- labor disputes, 461 ff.
- purposes and social philosophies, 439 ff.
- restraint of trade and, 497 ff.
- restrictions on entry into union, 467 f.
- right to exist, 490
- wage policies and business cycles, 678, 698 ff.

Transactions, volume of, 625 f.

Transfer agent, government as, 724 f.

Transportation, as marketing function of middleman, 33

Transportation facilities, controlling, 367 f.

Trusts:

- development of, 352 f.
- extreme price policies, examples of, 380 ff.

U**Uncertainty:**

- economic contribution and, 563 f.
- risk versus, 551 f.

Unemployment:

- borrowing in periods of, 796 f.
- expenditures in periods of, 804 f.
- new money in periods of, 800 f.
- taxation in periods of, 792
- see also* Business cycles

Unemployment insurance:

- benefits of, 817 f.
- coverage of, 816 f.
- economic aspects of, 815 ff.
- financing of, 817
- legal status of, 812 ff.
- nature of, 814 f.

United States:

- income distribution in, 525 ff.
- socialist ideology, compared with attitudes in, 565 f.
- sources of personal money incomes in, 555 ff.
- tariff making in, 871 ff.
- trade-unionism in:
 - checks to development of, 446 ff.
 - history of, 448 ff.
- types of money in, 582

United States Housing Act, regulation of wages, 509

Utilities, public, *see* Public utilities

Utility holding companies, regulation of, 426 ff.

V**Variable costs:**

- defined, 122
- position of firm and, in short run, 159

Variable unit cost:

- defined, 124
- relation between marginal unit costs and, 200

Variety, paying for, 331 f.

W**Wages:**

- base rate, determining, 514 ff.
- in business cycles, 667, 678, 698 ff.
- in purely competitive general equilibrium, 545 f.

minimum:

- difficulties in setting, 514 ff.
- effects of, 510 ff.
- legislation, history of, 506 ff.
- monopolistic and monopsonistic elements in determination of, 552
- policies, economic effects of, in trade-union bargaining, 471 ff.
- provisions in collective agreements, 460 f.

Wages and Hours Law, 509 f.

Walsh-Healy Act, regulation of wages in, 509

Wartime:

- "business cycle" policy in, 703 f.
- credit policy, federal reserve system, 651 f.
- expenditures, government, 727 ff.
- inflation, *see* Inflation; Price ceilings, and Fiscal policy.

Western Federation of Miners, 450

Withdrawal of firms, total cost and, 131 f.

Workers:

- attitudes and arguments of, in collective bargaining, 456 f.
- preferences of, realization, 482 f.
- union, resistance to displacement of, 470

Workmen's compensation, 821 ff.

Y

Yellow-dog contract, defined, 464

